



OPEN ACCESS

EDITED BY

Jianjun Ou,
The Second Xiangya Hospital of Central
South University, China

REVIEWED BY

Esther Stalujanis,
University Hospital Zürich, Switzerland
William Sulis,
McMaster University, Canada

*CORRESPONDENCE

Mohammad M. Herzallah
✉ mohammad.m.herzallah@alquds.edu

RECEIVED 21 May 2024

ACCEPTED 07 October 2024

PUBLISHED 05 December 2024

CITATION

Sawalma AS, Sehwal MA, Dammers J and
Herzallah MM (2024) Data-driven vs.
psychological personality temperaments:
theoretical and clinical utility of personality
measures in psychiatry.
Front. Psychiatry 15:1436121.
doi: 10.3389/fpsyt.2024.1436121

COPYRIGHT

© 2024 Sawalma, Sehwal, Dammers and
Herzallah. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Data-driven vs. psychological personality temperaments: theoretical and clinical utility of personality measures in psychiatry

Abdelrahman S. Sawalma^{1,2,3}, Mahmud A. Sehwal¹,
Jürgen Dammers^{2,3} and Mohammad M. Herzallah^{1,4*}

¹Palestinian Neuroscience Initiative, Al-Quds University, Jerusalem, Palestine, ²Institute of Neuroscience and Medicine (INM-4), Forschungszentrum Jülich, Jülich, Germany, ³Faculty of Medicine, Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen University, Aachen, Germany, ⁴Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, NJ, United States

Decades of research on personality identified dissociable psychological temperaments. Cloninger's temperament and character theory used a psychobiological approach to differentiate three major dimensions of personality: harm avoidance, novelty seeking, and reward dependence. Previous studies, heretofore, did not examine the correspondence between Cloninger's psychological temperaments and statistically independent data-driven components and how that could enhance the clinical utility of personality temperaments. In this study, we validated an Arabic version of the tri-dimensional personality questionnaire (TPQ) to construct data-driven personality temperaments using independent component analysis (ICA). Using SVM, we contrasted the clinical utility of data-driven personality vs. Cloninger's psychological temperaments in differentiating medication-naïve patients with major depressive disorder (N=244) and healthy subjects (N=1109). Data-driven personality components based on ICA showed very little overlap with Cloninger's original temperaments. Both Cloninger's temperaments and data-driven components revealed low internal consistency (for subscales) but high test-retest reliability. Cloninger's temperaments, however, showed a poor goodness-of-fit for the structure of the TPQ. Data-driven components significantly outperformed psychological TPQ temperaments with higher accuracy and recall but not precision. To our knowledge, this is the first study to examine the clinical utility of data-driven vs. psychological personality metrics using a sizeable sample of patients and healthy individuals. Our results could have wide implications for reexamining psychometric data to extract data-driven latent structures that can improve replicability, clinical utility, and cross-disciplinary inference.

KEYWORDS

personality, temperaments, dimension, temperament and character theory, tri-dimensional personality questionnaire, independent component analysis, major depressive disorder

Introduction

A myriad of psychological models explored the principal dimensions of personality. Collectively, these models varied in the number of proposed psychological dimensions of personality based on different survey methods, including the three-factor model (1, 2), the big-five model (3, 4), Strelau's temperament inventory (5), Catell's 16 personality factor model (6), the alternative model of personality disorders (AMPD) (7), and the Norman five (8). In the past three decades, Cloninger's psychobiological model of personality (9, 10) inferred significant impact on the field of psychiatry (11). Cloninger's model is based on neurochemical, neuroanatomical, and neurogenetic evidence, especially linking to monoaminergic activity in the brain. Thus, Cloninger's model presents a more suitable tool to study personality temperaments across psychiatric disorders (9, 10).

According to Cloninger, personality is expressed in three temperaments: novelty seeking (NS), representing exploration and impulsive actions; harm avoidance (HA), defined as cautiousness and fear of negative outcomes; and reward dependence (RD), manifested as seeking and maintenance of actions that result in positive outcomes (10). Each of these temperaments is further divided into four subscales according to different sub-temperaments. Cloninger used this categorization to create the tri-dimensional personality questionnaire (TPQ; 100 true/false questions) (10). Both English and translated versions of the TPQ revealed uniformly acceptable internal consistency and cross-cultural similarity [English (12), French (13), Japanese (14) and German (15, 16)]. It remains elusive, however, to what extent the data structure of question/item responses in the TPQ conform to the different temperaments put forward by Cloninger.

Previous studies utilized exploratory factor analysis (EFA) to examine the uncorrelated data components based on subscales and verify their correspondence with Cloninger's personality temperaments in the TPQ (15–19). Most studies reported results that are largely consistent with Cloninger's model (15, 16). Nevertheless, factors loadings were not unique to one subscale only, and some of them exhibited considerable cross-loadings on more than one subscale. Similarly, confirmatory factor analysis (CFA) results were inconclusive to support personality temperaments proposed by Cloninger (19–21). This is further corroborated by internal consistency measures (Cronbach's alpha) to test the structure of TPQ. At the temperament level, measures of internal consistency were within acceptable range (12). However, many of the subscales showed low internal consistency values (12, 19). Taken together, these lines of evidence show that the TPQ temperament and sub-temperament structure is not fully supported by data-driven constructs.

Various proposals appeared to restructure TPQ based on its questions/items using EFA (14, 22–24). Collectively, these studies failed to replicate Cloninger's structure and their internal consistency remained low. Moreover, EFA comes with multiple limitations related to the assumption of normality (Gaussianity) (25, 26) which cannot be expected in the TPQ where the answers are dichotomous (true or false). Further, in case of non-Gaussian data, EFA will only result in decorrelation of the signal and not

statistically-independent components (27). Independent component analysis (ICA), on the other hand, could decompose data into statistically independent components with no prior information about the nature of the underlying sources (28, 29). Statistical independence in ICA is computed using higher order moments, which are stronger statistical properties when compared to decorrelation (30). This warrants the use of ICA to identify the independent, rather than the uncorrelated, components when trying to find the personality components underlying the TPQ.

Personality components and temperaments were heavily studied in relation to psychiatric disorders (31). Compared to other models of personality, Cloninger's model in the TPQ showed significant promise in quantifying personality changes in psychiatric disorders. For instance, the TPQ showed higher HA in generalized anxiety disorder (32), panic disorder (32), eating disorders (33), obsessive compulsive disorder (34), following exposure to psychological trauma (35), and as a predictor of development of post-traumatic stress disorder (PTSD) (36). Conversely, patients with PTSD, alcoholism, and substance abuse exhibited higher levels of NS (16, 37). The correlation of TPQ temperaments and major depressive disorder (MDD) has received special attention. TPQ is characterized by high HA scores in patients with MDD (38, 39), as well as in healthy siblings of patients with MDD (40). Furthermore, TPQ temperaments, especially HA, were predictive of response to treatment in MDD (41, 42). However, significant overlaps are present within psychiatric disorders and within personality dimensions, which makes it difficult to discern independent personality dimensions when studying the clinical utility of TPQ, as evident in previous studies.

In general, a critical gap in studies of psychiatric disorders lies in the dissociation between the effects of these disorders and those stemming from exposure to psychotropic medications. The same applies to previous studies of TPQ in psychiatric populations. Previous studies recruited and mixed patients on-medications, patients off-medications, and medication-naïve patients. The lack of dissociation between disorder and medication effects could introduce multiple confounding variables (43, 44). Effects of chronicity (45) combined with the immediate vs. long-term effects of treatment (46) can also mask the personality dimensions of psychiatric disorder. This is further compounded by the vast comorbidity between different psychiatric disorders (47).

In this study, we used ICA to analyze TPQ at the question/item level to explore independent components based on a data-driven personality construct. We illustrated the differences between the data-driven and psychological personality temperaments as proposed by Cloninger. Further, we assessed the clinical utility of data-driven vs. psychologically-defined personality temperaments in patients with MDD. In particular, we attempted to circumvent previous limitations of studies on psychiatric populations by testing medication-naïve patients with MDD. For all aforementioned analyses, we used EFA to analyze TPQ results as a control condition. Compared to Cloninger's temperaments and EFA results, we predicted that our data-driven temperaments will be principally different and possess diagnostic and treatment-predictive utilities that outperform the psychological temperaments.

Methods

Protocols and consent

All subjects provided written informed consent before enrollment in the study. Study protocols were in accordance with the Declaration of Helsinki and approved the Al-Quds University Research Ethics Committee.

Subjects

We tested 1353 Arabic-speaking subjects in the West Bank, Palestine, including 244 patients with MDD and 1109 healthy controls (HC). A subsample of 90 HC was tested and retested 4-6 weeks apart to check for the test-retest reliability of the questionnaire. Participants were recruited from Al-Quds University, An-Najah University, and various psychiatric and neurological clinics throughout the West Bank, Palestine.

Exclusion criteria for all subjects included prior psychotropic drug exposure; the presence of any psychiatric disorder except for MDD, major medical or neurological illness, illicit drug use or alcohol abuse within the past year, lifetime history of alcohol or drug dependence and current pregnancy or breastfeeding.

Aside from completing an Arabic-translated version of the TPQ (48), all subjects were interviewed using the mini-international neuropsychiatric interview (MINI) (49), the Beck depression inventory II (BDI-II) (50) and the Beck anxiety inventory (BAI) (51). Table 1 summarizes demographics and neuropsychological characteristics for our sample.

TPQ validation

For initial validation of the translation, the TPQ was translated to Arabic and back-translated to English for cross-referencing. Following data collection, we used HC data for validation of our Arabic translation of the TPQ. In particular, we calculated Cronbach's alpha for internal consistency estimation for TPQ scales and subscales. Further, we performed a test-retest reliability on a subset of 90 subjects using Spearman's rho. Internal consistency measures for the main scales were within the acceptable range for the

three main scales, ranging from 0.63 to 0.82. Cronbach alpha scores were acceptable for HA subscales and poor for NS and RD subscales, except for RD3, which had acceptable score of 0.70. Test-retest reliability was moderately high for the main scales, ranging from 0.70 to 0.81, and moderate to high in the subscales, ranging from 0.42 to 0.8. Table 2 shows the descriptive statistics for TPQ scales and subscales along with Cronbach alpha values. Our validation results were similar to previous validation and normative studies of the TPQ (12, 15, 16, 19).

Exploratory & confirmatory factor analysis for TPQ subscales

We performed exploratory factor analysis (EFA) on the 12 scales of the TPQ using a three-factor solution with a varimax rotation (52). We followed EFA with confirmatory factor analysis (CFA) using the *lavaan* R package to measure the resemblance between the resulting factors and the original Cloninger scales (53). CFA included measures of goodness of fit: p-value, comparative fit index (CFI), root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR). We used the following criteria to for goodness of fit: a non-significant chi-square p-value (20), CFI > 0.90 (54), RMSEA < 0.08 (54) and SRMR <0.10 (54).

Independent component analysis

We applied ICA at the item/question level to extract independent components of personality in our volunteers. In particular, we used Infomax ICA which utilizes an unsupervised neural network for maximizing the joint entropy to find the independent components (28, 55, 56). Prior to applying ICA, principal component analysis (PCA) was applied for data whitening and dimension reduction (26). Using parallel analysis, we were able to determine the appropriate number of PCA components to retain post whitening. In parallel analysis, we created a simulated dataset from which we calculated the eigenvalues. The number of components is determined as the number of original data components with eigenvalues higher than the intersection point between simulated data components (57). In our dataset, using parallel analysis indicated that the number of components that pass the parallel analysis threshold in our dataset was 13 (from 100). We wrote the code for parallel analysis in Python based on the R package *psych* (58). Parallel analysis results are illustrated in Figure 1. Accordingly, our subsequent ICA was expected to produce 13 independent components (ICs) based on the TPQ data.

Support vector machine

We used SVM to compare the ability of ICA data-driven components vs. Cloninger's original TPQ temperaments in differentiating patients with MDD from HC. The ICA dataset contained 13 independent components extracted from TPQ

TABLE 1 Demographics and basic psychometric measures of subjects in both the HC and MDD groups. Age and Education are given in years and are presented as mean ± standard deviation.

	HC	MDD
Sample Size	1109	244
Age	22.3 ± 9.2	28.6 ± 9.7
Education	14.0 ± 3.6	12.7 ± 3.0
Female Percentage	59.6	60.7
BDI score	10.1	30.7
BAI score	10.6	28.3

TABLE 2 Descriptive statistics and internal consistency measures for HC, divided into males and females.

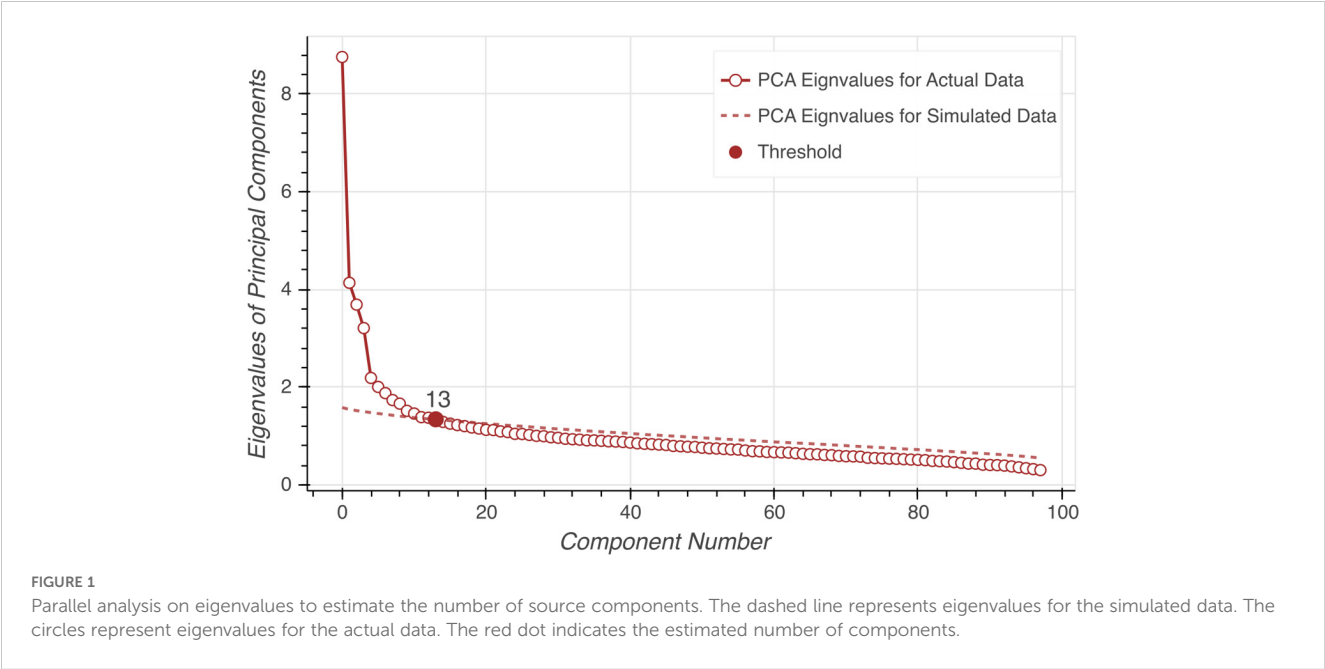
TPQ Measure	Mean All (mean ± SD)	Mean Male (mean ± SD)	Mean Female (mean ± SD)	Cron. α All	Cron. α Male	Cron. α Female	Test-retest Reliability
NS1	5.1 ± 1.6	5.0 ± 1.6	5.2 ± 1.6	0.35	0.35	0.35	0.57*
NS2	2.6 ± 1.7	2.6 ± 1.6	2.5 ± 1.8	0.50	0.45	0.54	0.58*
NS3	3.5 ± 1.6	3.5 ± 1.6	3.5 ± 1.6	0.55	0.53	0.58	0.81*
NS4	4.6 ± 1.9	4.6 ± 1.9	4.5 ± 1.8	0.39	0.42	0.36	0.73*
NS	15.7 ± 4.2	15.8 ± 4.1	15.8 ± 4.3	0.63	0.60	0.65	0.70*
HA1	3.5 ± 2.1	3.5 ± 2.1	3.5 ± 2.2	0.61	0.57	0.64	0.66*
HA2	4.0 ± 1.9	4.0 ± 1.9	4.0 ± 1.9	0.63	0.65	0.63	0.67*
HA3	1.9 ± 1.6	1.9 ± 1.7	1.9 ± 1.7	0.60	0.60	0.60	0.80*
HA4	3.6 ± 2.5	3.5 ± 2.5	3.7 ± 2.6	0.73	0.72	0.74	0.78*
HA	13.0 ± 5.9	12.8 ± 5.8	13.1 ± 6.1	0.82	0.81	0.83	0.81*
RD1	4.3 ± 0.9	4.3 ± 1.0	4.4 ± 0.9	0.41	0.45	0.40	0.56*
RD2	4.1 ± 1.7	4.3 ± 1.6	4.0 ± 1.8	0.38	0.29	0.43	0.42*
RD3	7.0 ± 2.5	6.8 ± 2.4	7.1 ± 2.5	0.70	0.68	0.72	0.72*
RD4	2.0 ± 1.3	1.9 ± 1.3	2.0 ± 1.3	0.45	0.41	0.47	0.63*
RD	17.4 ± 3.8	17.2 ± 3.7	17.5 ± 3.9	0.61	0.58	0.62	0.71*

SD, standard deviation *: significant Spearman’s rho at p < 0.001.

questions/items using Infomax ICA. The temperament subscale dataset separated TPQ results into 12 subscales according to the original approach suggested by Cloninger (10).

Our sample contained 1109 HC and 244 patients with MDD. We split the dataset using an 80/20 ratio for training and holdout. In order to overcome the limitations of the imbalanced dataset, we used a combination of undersampling of the majority group, and oversampling the minority group. For oversampling, we utilized

synthetic minority oversampling technique (SMOTE) (59) which introduces new examples in the minority class along the line segments joining the k nearest neighbors (59, 60). SMOTE was combined with undersampling, which is done by randomly removing samples from the majority group as implemented in the python package *imbalanced-learn* (61). Studies suggest that using SMOTE in combination with random undersampling provides better classification results as compared to undersampling only



(59). We applied the SMOTE/undersampling approach to the training and holdout datasets independently.

We applied SVM on the balanced training dataset to test the clinical utility of ICA data-driven vs. subscale variables vs TPQ factors in identifying patients with MDD (62). The data were divided into training, validation and holdout datasets with ratios 64/16/20. Optimal hyperparameters were estimated using grid search as implemented in Python *scikit-learn* (63). Grid search compares combinations of different types of kernels (e.g., linear, radial basis function (RBF) or sigmoid) and kernel function parameters (hyperplane confidence interval (C), linearity of separation (γ)) to find the optimal hyperparameters. The median hyperparameters are shown in Table 3. To prevent overfitting, SVM was run 5 times in a 5-fold cross validation scheme, where we changed the subjects assigned to the holdout set with every iteration. A workflow of the model training and testing is shown in Figure 2.

Results

EFA and CFA of TPQ

We used EFA to investigate whether uncorrelated factors of TPQ subscales conform to Cloninger's TPQ scales. Factor loadings of subscales following varimax rotation are shown in Table 4. For simplification, we removed all absolute factor loadings below 0.3. The three-factor solution largely replicates Cloninger's three temperaments, with some deviations in NS1, HA2 and RD2. The variance explained by the factor model was 33%.

TABLE 3 Best parameters chosen by grid search cross validation for each data set.

Hyperparameter	Subscales Dataset	ICA Dataset	FA Dataset
C	5	3	3
γ	0.1	0.1	25
Kernel	RBF	RBF	RBF

Grid search cross validation was applied 4 times, one for each training data set produced. The median of the parameters used are mentioned here. RBF, radial basis function.

We used CFA to confirm whether Cloninger's three temperament model of TPQ fits our data. CFA measures showed a poor goodness-of-fit. Three measures showed poor goodness-of-fit while the fourth was acceptable. CFA goodness-of-fit results can be found in Table 5.

Data-driven components of the TPQ using ICA

To complement our analysis on the original TPQ temperaments and subscales, we conducted both Cronbach's alpha and a test-retest reliability using Spearman's rho on the data-driven personality components. ICA produced 13 components that represent the data-driven TPQ structure. In order to project the components onto the question space, we assigned each component with the questions with the highest 10th percentile of loadings. Below is a summary of the internal consistency and test-retest reliability of the data-driven components (Table 6).

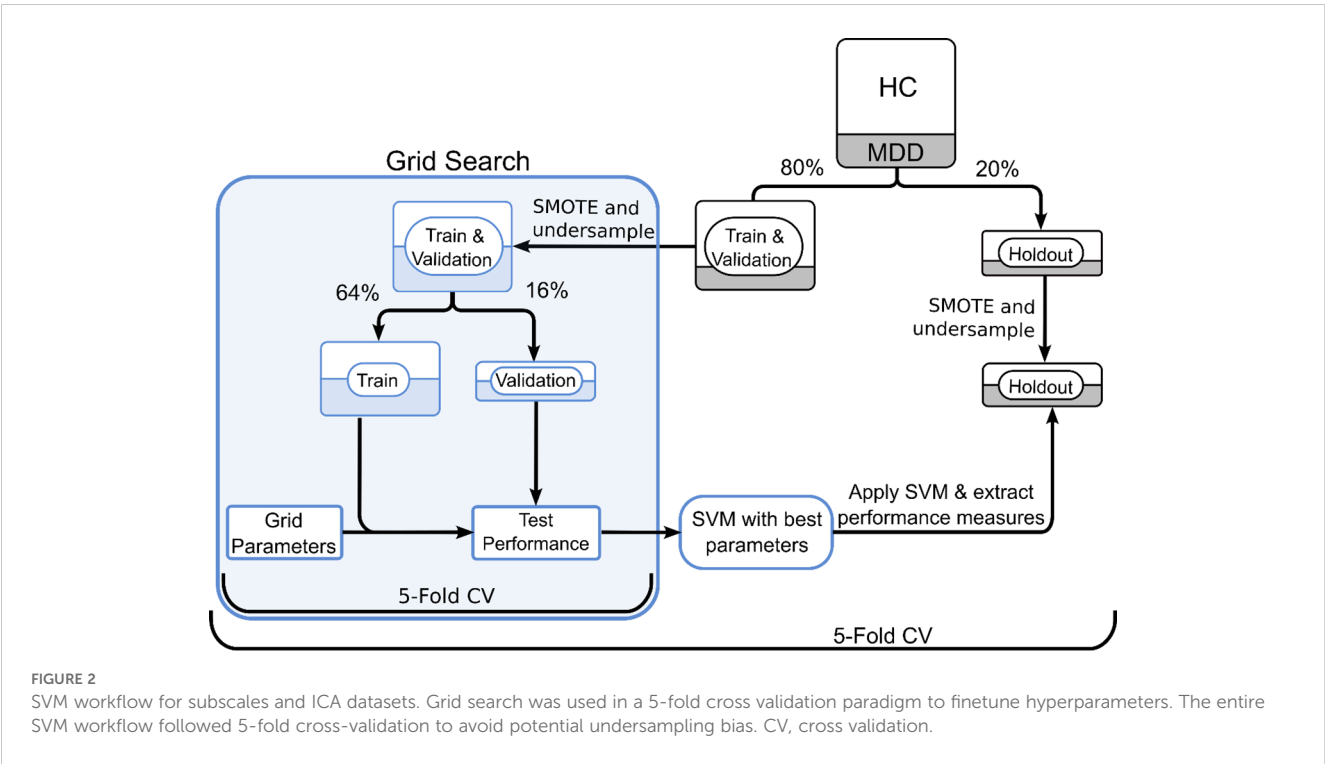


TABLE 4 Factor loadings onto TPQ subscales for a three-factor solution using EFA.

	F1	F2	F3
NS1	-0.37		
NS2		0.49	
NS3		0.34	
NS4		0.64	
HA1	0.74		
HA2	0.56	-0.34	
HA3	0.67		
HA4	0.7		
RD1			-0.35
RD2			
RD3			-0.68
RD4			-0.43
Cumulative Variance	0.17	0.26	0.33

Shown here are loadings with absolute value > 0.3.

In order to understand whether a correlation between the ICs and the TPQ subscales and factors, we identified the loading of the separate ICs on the Subscales and factors. For the subscales, we summed the ICA weights for the questions that represent each subscale. As for the TPQ factors, we summed the product of IC weights and factor weights for each IC and factor. The z-scored ICA loadings are represented in [Figure 3](#).

TABLE 5 Goodness-of-fit values for the CFA Model.

Measure	Value
χ^2	460.936
p-value	<0.001
CFI	0.744
RMSEA	0.085
SRMR	0.083

CFI, comparative fit index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual.

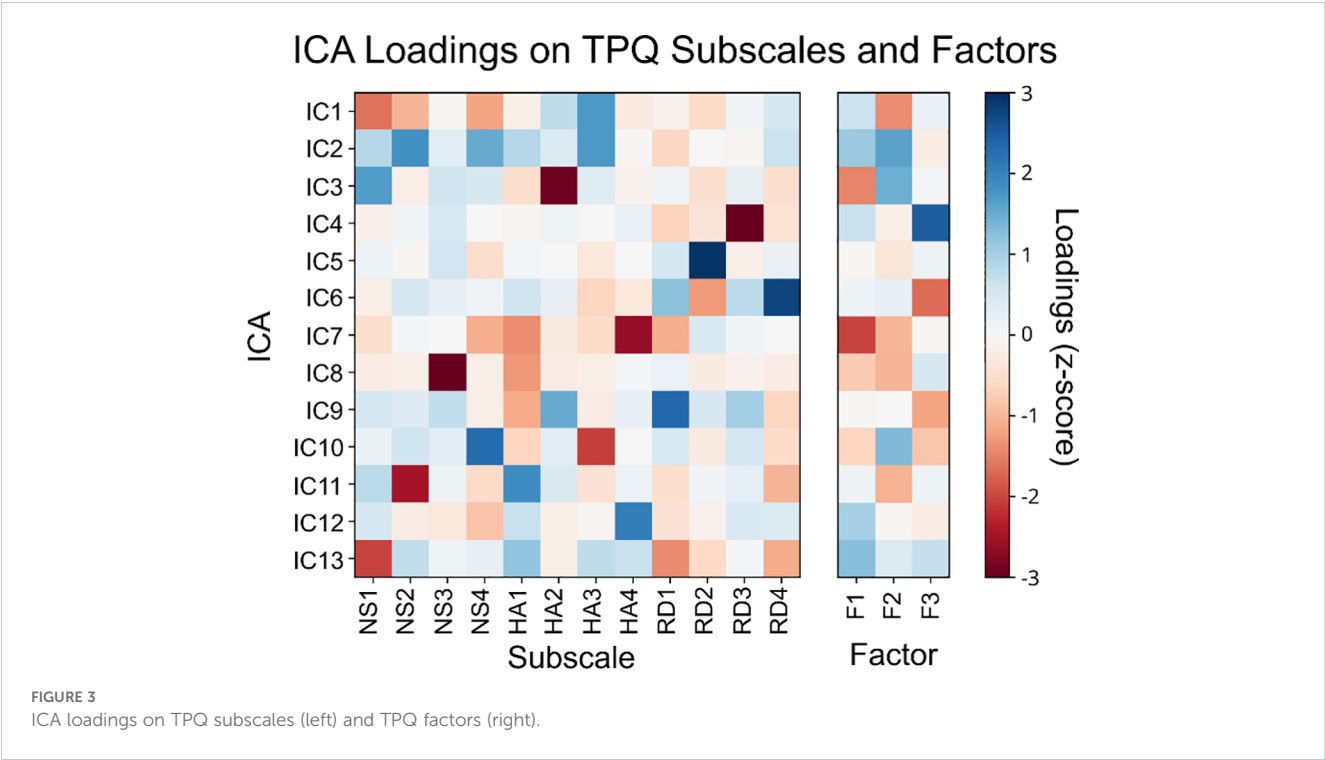
Identifying MDD using TPQ psychological vs. data-driven temperaments

We used the SVM-based classification workflow to compare data-driven ICA components to TPQ subscales and TPQ Factors for identifying MDD. Kruskal-Wallis Test was conducted to examine the differences in accuracy, precision, recall and AUC. The results show a significant difference in all four comparisons (Kruskal-Wallis H = 11.1, 9.78, 11.81 and 11.58, respectively, p-value <0.008 for all four comparisons). To follow up on this test, we performed Mann-Whitney’s test for the four comparisons with Holm-Bonferroni correction. For accuracy, recall and AUC comparisons, ICA-based SVM was higher than both the subscales- and the FA-based SVM, with corrected p-values ≤ 0.035 in all cases. As for precision, ICA-based SVM was not significantly different from Subscales-based SVM, and both were higher than FA-based SVM, with corrected p-values = 0.024 in both cases. Thus, our SVM analysis on ICA components outperformed the one based on Cloninger’s TPQ subscales in accuracy, recall and

TABLE 6 Descriptive statistics and internal consistency measures for the derived independent components.

	Mean All (mean ± SD)	Mean Male (mean ± SD)	Mean Female (mean ± SD)	Cron. α	Cron. α Male	Cron. α Female	Test- Retest Reliability
IC1	4.0 ± 2.5	4.1 ± 2.5	3.9 ± 2.5	0.71	0.72	0.71	0.619
IC2	5.2 ± 2.4	5.1 ± 2.3	5.3 ± 2.4	0.66	0.63	0.68	0.696
IC3	2.9 ± 1.8	3.0 ± 1.7	2.9 ± 1.8	0.41	0.32	0.47	0.606
IC4	4.8 ± 1.7	4.7 ± 1.6	5.0 ± 1.6	0.22	0.2	0.22	0.736
IC5	4.2 ± 2.0	4.3 ± 1.9	4.2 ± 2.1	0.46	0.39	0.51	0.583
IC6	5.6 ± 1.6	5.5 ± 1.6	5.7 ± 1.6	0.13	0.1	0.15	0.747
IC7	4.6 ± 2.3	4.4 ± 2.2	4.8 ± 2.4	0.65	0.61	0.67	0.623
IC8	4.4 ± 1.7	4.5 ± 1.8	4.2 ± 1.7	0.37	0.39	0.36	0.825
IC9	4.7 ± 1.7	4.5 ± 1.7	4.8 ± 1.7	0.29	0.25	0.32	0.654
IC10	4.5 ± 1.6	4.6 ± 1.7	4.5 ± 1.6	0.17	0.24	0.09	0.644
IC11	4.3 ± 1.7	4.3 ± 1.6	4.3 ± 1.7	0.2	0.14	0.25	0.579
IC12	4.8 ± 2.0	4.6 ± 2.0	5.0 ± 2.0	0.47	0.43	0.49	0.676
IC13	3.9 ± 1.7	3.9 ± 1.7	3.8 ± 1.7	0.2	0.19	0.21	0.503

Each component is represented by the 10 questions that have the highest weight. SD, standard deviation; *, significant Spearman’s rho at p < 0.001.



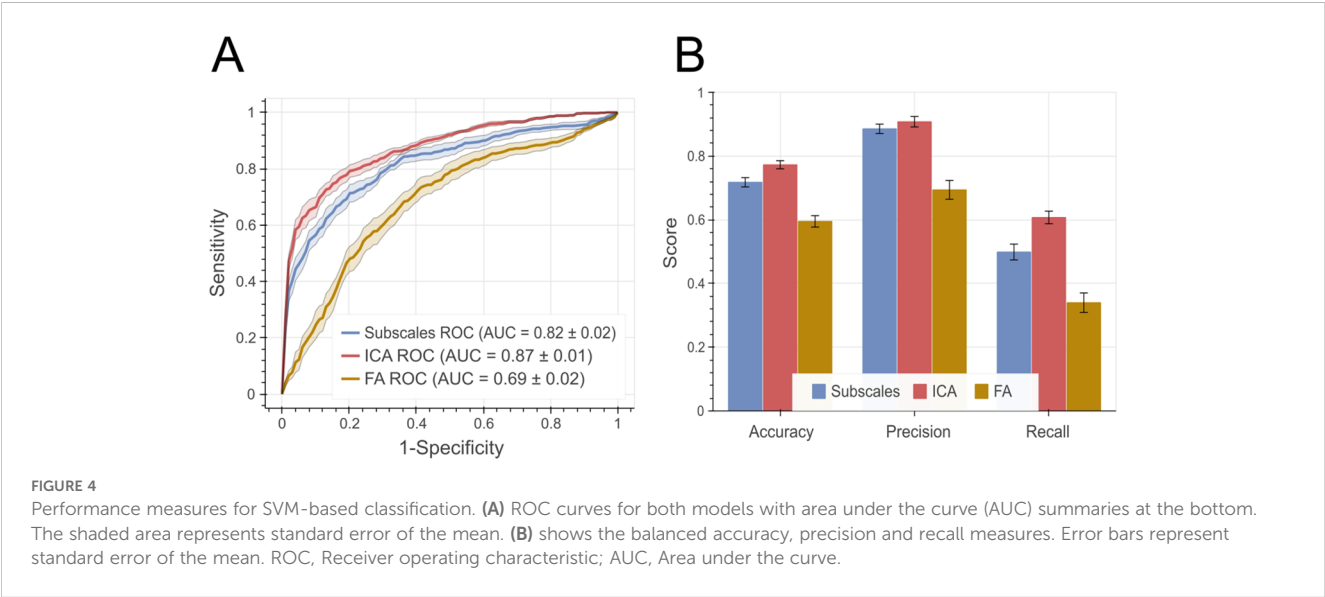
area under the curve (AUC), while precision was comparable, as summarized in [Figure 4](#).

Discussion

We investigated the data-driven vs. psychological structure of an Arabic-translation of Cloninger’s TPQ and its clinical utility. Our results revealed a discordance between data-driven and theory-driven classification of TPQ data. TPQ internal consistency and test-retest reliability results echoed previous findings by showing acceptable levels for the main three temperaments, but subscales did

not perform as well. Using EFA produced a three-factor solution with loadings close to Cloninger’s subscale proposal. However, CFA exhibited a poor goodness-of-fit for the three-factor solution. Data-driven personality components using ICA did not corroborate Cloninger’s classification. Similar to Cloninger’s subscales, data-driven personality components showed low internal consistency but high test-retest reliability. In its clinical utility, data-driven personality components showed significantly better accuracy and recall for identifying patients with MDD than Cloninger’s theory-driven personality temperaments.

This study utilized Cloninger’s model given its psychobiological basis, especially the link between personality temperaments to



various monoamines (9, 10). In contrast, other models of personality, such as the big five and its derivatives, assume that personality differences are lexically reflected in language (4, 64). Thus, their utility would better fit socio-cultural rather than psychobiological studies (65). The structure of temperament questionnaire (STQ) covers behaviorally-based components without sufficient inference to basic biological processes (66). More recent developments such as the functional ensemble of temperament (FET) were conceived after the initiation of data collection for the current study (67). Other clinically-oriented models, like the AMPD (included in the diagnostic and statistical manual of mental disorders (DSM-5)) (7), provides clinical evaluation of personality disorders (pathological personality), as opposed to personality temperaments in psychiatric disorders (68). With neurochemical, neuroanatomical, and neurogenetic basis, Cloninger's model presented the most befitting choice to investigate personality temperaments in MDD (9, 10).

Our internal consistency results corroborated previously published normative data in English and translated versions of the TPQ (12, 15, 16, 19). Specifically, major temperaments showed high (for HA) or acceptable (NS and RD) internal consistency. In contrast, internal consistency was acceptable for HA subscales but low for NS and RD subscales. Our data show similar results for HA main scale and subscales, but lower internal consistency for RD and NS scales and subscales, falling below 0.60 in most subscales. Test-retest reliability was moderately high, confirming the reliability of TPQ when applied at different time points, which falls in line with previous findings as well (10, 12). Some of the inconsistencies with previous findings could be attributed to our significantly larger sample size. Further, we used stricter inclusion criteria for HC which could have unraveled previously unobserved characteristics of TPQ metrics.

We tested whether Cloninger's three-temperaments of the TPQ using both EFA and CFA. Overall, EFA produced a three-factor solution close to that proposed by Cloninger with some deviations. In line with previous studies, the HA2 subscale showed high loading on the EFA factor for NS (16, 19) while RD2 did not have loadings on any of the three EFA factors (18). For validation of the goodness-of-fit of the three-factor solution of TPQ, we utilized CFA to test the extent to which the data fits the proposed variable structure (20). We applied CFA on the subscale level to test if the main scales can be reliably derived from the subscales. Our results revealed a poor goodness of fit for Cloninger's model, consistent with previous findings (19, 69). Surprisingly, previous examination of CFA results on the TPQ overlooked the poor goodness-of-fit results and interpreted other metrics, such as the goodness of fit and root mean square, in isolation (20, 21). Taken together, our EFA and CFA results and previous findings indicated that there is more to the data-driven structure of the TPQ than what Cloninger's model provides.

With ICA, we utilized more objective, non-Gaussian, and assumption-free approach to unravel the data-driven structure of TPQ. Previous methods to describe the structure of the TPQ were either subjective, i.e., the researcher deciding which questions belong to which scale based on *a priori* theoretical accounts (10), or partially data-driven using factor analysis on predefined subscales that are

based on *a priori* theoretical accounts (14, 22–24). Clearly, subjective categorization of TPQ results was not optimal given the low internal consistency at the subscale level and the poor goodness-of-fit (12, 15, 16, 19). On the other hand, given the binomial distribution of responses to TPQ questions/items (true/false), factor analysis of the TPQ data can lead to biased conclusions given its assumption on the Gaussianity of data (25, 26). Nevertheless, previous studies used factor analysis at the question/item level and could not replicate Cloninger's TPQ structure (14, 22–24). In comparison, ICA does not assume data normality (70), as evident in TPQ data. Furthermore, ICA defines statistical independence using higher order moments as opposed to decorrelation with factor analysis (30). Application of ICA produced 13 independent components that represent the data-driven TPQ structure. ICA components were drastically different from the 12 subscales of Cloninger's TPQ structure as well as the 3-factor solution of EFA (10, 12).

Our data-driven components exhibited consistently moderate to high test-retest reliability was across all components, thus confirming the reliability of the data-driven TPQ structure. However, the internal consistency was poor overall, with the exception of IC1, IC2, and IC7, which were deemed acceptable. This is likely due to the fact that the independent components represent statistically independent, and mostly nonlinear, latent variables with weights derived from all questions rather than from a subset of questions only. As such, ICA loadings did not uniquely relate to any specific subset of the TPQ questions (or Cloninger's subscales or EFA factors). As such, our results confirmed the difference between data-driven personality components and the 12 TPQ subscales proposed by Cloninger (10, 12).

TPQ has been widely used to study psychopathology and psychotropic treatment effects. Here, however, we used TPQ metrics as potential diagnostic markers for MDD. We further compared the clinical utility of the data-driven vs. the psychological structure of the TPQ in identifying MDD in a sizeable sample of patients and healthy individuals. The data-driven components of TPQ were significantly more superior to the psychological structure of the TPQ in distinguishing medication-naïve patients with MDD from HC. Specifically, both accuracy (identification of MDD and HCs) and recall (identification of MDD) were more pronounced for the data-driven classifier. This might be attributed to the nonlinear and non-Gaussian examination the data compared to factor analysis, or factor-analysis-based categorization methods. These methods are often criticized as being simple, and unable to capture the complexity of underlying psychological processes (65). This result holds significant potential for data-driven reexamination of psychometric results to extract statistically independent components that link better to underlying behavioral and neural constructs or subsequent diagnosis and treatment choices (71).

Finally, we fully recognize that identifying statistically-independent components does not necessarily equate with measuring underlying biological processes of temperaments or personality characteristics, which are often more complex and interdependent (65). For instance, the functional ensemble of temperament (FET) model highlights the role neurotransmitter system in the regulation of temperaments, with their items

reflecting the dynamics of behavior (31, 66, 72). Although data collection for our study predates the neurochemically-informed FET, examination of results from the FET model using our approach would shed a very important light on the correspondence between data-driven temperaments and their neurochemical basis. In fact, we argue that every study using psychometrics should cross-examine the data-driven and theory-based structure of their results to highlight latent variables that could link different levels of analysis.

To our knowledge, this is the first study that combines advanced analysis of TPQ structure, the utilization of machine learning classifiers to identify MDD, and a significant sample size of both patients and matched controls. In essence, this paper presents a roadmap for the reexamination of a huge body of literature on psychometrics. Immediate positive outcomes for the field of mental health can emerge in the form of improving clinical utility, replicability, and cross-discipline inferences.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Al-Quds University Research Ethics Committee Al-Quds University Jerusalem Palestine. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

AS: Data curation, Formal analysis, Software, Visualization, Writing – original draft, Writing – review & editing. MS: Data curation, Project administration, Supervision, Writing – review &

editing. JD: Conceptualization, Formal analysis, Supervision, Validation, Writing – review & editing. MH: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was sponsored as part of research programs at the Palestinian Neuroscience Initiative by the Arab Palestinian Investment Company (APIC), Hikma Pharmaceuticals LLC., and GMS Holdings, while the German Federal Ministry of Education and Research (BMBF; through the Palestinian German Science Bridge and grant number 01DH16027) sponsored AS's graduate studies at the FZJ. The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Eysenck HJ. *The biological basis of personality*. Springfield, IL; Charles Thomas (1967).
2. Eysenck HJ. *The structure of human personality*. London, UK: Methuen (1970).
3. McCrae RR, Costa PT. Validation of the five-factor model of personality across instruments and observers. *J Pers Soc Psychol*. (1987) 52(1):81.
4. Costa PT, McCrae RR. Normal personality assessment in clinical practice: the NEO personality inventory. *Psychol Assess*. (1992) 4:5–13. doi: 10.1037/1040-3590.4.1.5
5. Strelau J, Angleitner A, Bantelmann J, Ruch W. The Strelau Temperament Inventory-revised (STI-R): Theoretical considerations and scale development. *Eur J Pers*. (1990) 4:209–35. doi: 10.1002/per.2410040304
6. Cattell RB, Eber HW, Tatsuoka MM. *Handbook for the Sixteen Personality Factor Questionnaire (16PF)*. Champaign, IL: Institute for Personality and Ability Testing (1970).
7. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-5*. Washington, DC: American psychiatric association (2013).
8. Norman WT. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *J Abnorm Soc Psychol*. (1963) 66:574–83. doi: 10.1037/h0040291
9. Cloninger CR. A unified biosocial theory of personality and its role in the development of anxiety states. *Psychiatr Dev*. (1986) 4:167–226.
10. Cloninger CR. A systematic method for clinical description and classification of personality variants. *Arch Gen Psychiatry*. (1987) 44:573. doi: 10.1001/archpsyc.1987.01800180093014
11. de Fruyt F, van de Wiele L, van Heeringen C, de Fruyt F, van de Wiele L, van Heeringen C. Cloninger's psychobiological model of temperament and character and the five-factor model of personality. *Pers Individ Dif*. (2000) 29:441–52. doi: 10.1016/S0191-8869(99)00204-4
12. Cloninger CR, Przybeck TR, Svrakic DM. The Tridimensional Personality Questionnaire: U.S. normative data. *Psychol Rep*. (1991) 69:1047–57. doi: 10.2466/pr0.1991.69.3.1047
13. Le Bon O, Staner L, Tecco J, Pull C, Pelc I. Tridimensional Personality Questionnaire (TPQ): validation in a French-speaking control population. *Encephale*. (1998) 24:40–5.
14. Takeuchi M, Yoshino A, Kato M, Ono Y, Kitamura T. Reliability and validity of the Japanese version of the tridimensional personality questionnaire among university students. *Compr Psychiatry*. (1993) 34:273–9. doi: 10.1016/0010-440X(93)90010-2

15. Weyers P, Krebs H, Janke W. Reliability and construct validity of the German version of Cloninger's Tridimensional Personality Questionnaire. *Pers Individ Dif.* (1995) 19:853–61. doi: 10.1016/S0191-8869(95)00128-X
16. Krebs H, Weyers P, Janke W. Validation of the German version of Cloninger's TPQ: Replication and correlations with stress coping, mood measures and drug use. *Pers Individ Dif.* (1998) 24:805–14. doi: 10.1016/S0191-8869(98)00008-7
17. Stallings MC, Hewitt JK, Cloninger CR, Heath AC, Eaves LJ. Genetic and environmental structure of the Tridimensional Personality Questionnaire: three or four temperament dimensions? *J Pers Soc Psychol.* (1996) 70:127–40. doi: 10.1037/0022-3514.70.1.127
18. Kleinfeld EI, Sunday S, Hurt S, Halmi KA. Psychometric validation of the tridimensional structure of the Tridimensional Personality Questionnaire: Application to subgroups of eating disorders. *Compr Psychiatry.* (1993) 34:249–53. doi: 10.1016/0010-440X(93)90006-P
19. Otter C, Huber J, Bonner A. Cloninger's Tridimensional Personality Questionnaire: Reliability in an english sample. *Pers Individ Dif.* (1995) 18:471–80. doi: 10.1016/0191-8869(94)00199-3
20. Bagby RM, Parker JDA, Joffe RT. Confirmatory factor analysis of the Tridimensional Personality Questionnaire. *Pers Individ Dif.* (1992) 13:1245–6. doi: 10.1016/0191-8869(92)90260-V
21. Parker JDA, Bagby RM, Joffe RT. Validation of the biosocial model of personality: confirmatory factor analysis of the tridimensional personality questionnaire. *Psychol Assess.* (1996) 8:139–44. doi: 10.1037/1040-3590.8.2.139
22. Stewart ME, Ebmeier KP, Deary IJ. The structure of Cloninger's Tridimensional Personality Questionnaire in a British sample. *Pers Individ Dif.* (2004) 36:1403–18. doi: 10.1016/S0191-8869(03)00237-X
23. Zohar AH, Lev-Ari L, Benjamin J, Ebstein R, Lichtenberg P, Osher Y. The psychometric properties of the Hebrew version of Cloninger's Tridimensional Personality Questionnaire. *Pers Individ Dif.* (2001) 30:1297–309. doi: 10.1016/S0191-8869(00)00112-4
24. Cannon DS, Clark LA, Leeka JK, Keefe CK. A reanalysis of the tridimensional personality questionnaire (TPQ) and its relation to cloninger's type 2 alcoholism. *Psychol Assess.* (1993) 5:62–6. doi: 10.1037/1040-3590.5.1.62
25. Wirth RJ, Edwards MC. Item factor analysis: current approaches and future directions. *Psychol Methods.* (2007) 12:58. doi: 10.1037/1082-989X.12.1.58
26. Gie Yong A, Pearce S. A beginner's guide to factor analysis: focusing on exploratory factor analysis. *Quant Methods Psychol.* (2013) 9:79–94. doi: 10.20982/tqmp.09.2.p079
27. Hyvärinen A, Oja E. Independent component analysis: Algorithms and applications. *Neural Networks.* (2000) 13:411–30. doi: 10.1016/S0893-6080(00)00026-5
28. Bell AJ, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* (1995) 7:1129–59. doi: 10.1162/neco.1995.7.6.1129
29. Hyvärinen A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Netw.* (1999) 10:626–34. doi: 10.1109/72.761722
30. Verma PK, Gaikwad AN, Singh D, Nigam MJ. Analysis of clutter reduction techniques for through wall imaging in uwb range. *Prog In Electromagnetics Res B.* (2009) 17:29–48. doi: 10.2528/pierb09060903
31. Trofimova I, Sulis W. There is more to mental illness than negative affect: Comprehensive temperament profiles in depression and generalized anxiety. *BMC Psychiatry.* (2018) 18:1–17. doi: 10.1186/s12888-018-1695-x
32. Starcevic V, Uhlenhuth EH, Fallon S, Pathak D. Personality dimensions in panic disorder and generalized anxiety disorder. *J Affect Disord.* (1996) 37:75–9. doi: 10.1016/0165-0327(95)00058-5
33. Brewerton TD, Hand LD, Bishop ER. The tridimensional personality questionnaire in eating disorder patients. *Int J Eating Disord.* (1993) 14:213–8. doi: 10.1002/1098-108X(199309)14:2<213::AID-EAT2260140211>3.0.CO;2-P
34. Pfohl B, Black D, Noyes R, Kelley M, Blum N. A test of the tridimensional personality theory: association with diagnosis and platelet imipramine binding in obsessive-compulsive disorder. *Biol Psychiatry.* (1990) 28:41–6. doi: 10.1016/0006-3223(90)90430-A
35. Shin SM, Lee BW, Yi JS. Comparison of adolescents' Temperament and character inventory (TCI) profile between traumatized group and non-traumatized group, and between PTSD symptom group and non-PTSD symptom group among the traumatized group. *Anxiety Mood.* (2012) 8:133–40.
36. Gil S. Pre-traumatic personality as a predictor of post-traumatic stress disorder among undergraduate students exposed to a terrorist attack: A prospective study in Israel. *Pers Individ Dif.* (2005) 39:819–27. doi: 10.1016/j.paid.2005.03.007
37. Wang S, Mason J, Charney D, Yehuda R, Riney S, Southwick S. Relationships between hormonal profile and novelty seeking in combat-related posttraumatic stress disorder. *Biol Psychiatry.* (1997) 41:145–51. doi: 10.1016/S0006-3223(95)00648-6
38. Hansenne M, Pitchot W, Gonzalez Moreno A, Reggers J, Machurot PY, Ansseau M. Harm avoidance dimension of the tridimensional personality questionnaire and serotonin-1a activity in depressed patients. *Biol Psychiatry.* (1997) 42:959–61. doi: 10.1016/S0006-3223(97)00325-9
39. Nelson E, Cloninger CR. Exploring the TPQ as a possible predictor of antidepressant response to nefazodone in a large multi-site study. *J Affect Disord.* (1997) 44:197–200. doi: 10.1016/S0165-0327(97)00047-5
40. Farmer A, Mahmood A, Redman K, Harris T, Sadler S, McGuffin P. A sib-pair study of the Temperament and Character Inventory scales in major depression. *Arch Gen Psychiatry.* (2003) 60:490–6. doi: 10.1001/archpsyc.60.5.490
41. Joyce PR, Mulder RT, Cloninger CR. Temperament predicts clomipramine and desipramine response in major depression. *J Affect Disord.* (1994) 30:35–46. doi: 10.1016/0165-0327(94)90149-X
42. Joyce PR, McKenzie JM, Carter JD, Rae AM, Luty SE, Frampton CM, et al. Temperament, character and personality disorders as predictors of response to interpersonal psychotherapy and cognitive-behavioural therapy for depression. *Br J Psychiatry.* (2007) 190:503–8. doi: 10.1192/bjp.bp.106.024737
43. Herzallah MM, Moustafa AA, Natsheh JY, Abdellatif SM, Taha MB, Tayem YI, et al. Learning from negative feedback in patients with major depressive disorder is attenuated by SSRI antidepressants. *Front Integr Neurosci.* (2013) 7:67. doi: 10.3389/fnint.2013.00067
44. Herzallah MM, Moustafa AA, Natsheh JY, Danoun OA, Simon JR, Tayem YI, et al. Depression impairs learning, whereas the selective serotonin reuptake inhibitor, paroxetine, impairs generalization in patients with major depressive disorder. *J Affect Disord.* (2013) 151:484–92. doi: 10.1016/j.jad.2013.06.030
45. MacQueen GM, Galway TM, Hay J, Young LT, Joffe RT. Recollection memory deficits in patients with major depressive disorder predicted by past depressions but not current mood state or treatment status. *Psychol Med.* (2002) 32:251–8. doi: 10.1017/S0033291701004834
46. Harmer CJ, Goodwin GM, Cowen PJ. Why do antidepressants take so long to work? A cognitive neuropsychological model of antidepressant drug action. *Br J Psychiatry.* (2009) 195:102–8. doi: 10.1192/bjp.bp.108.051193
47. van Loo HM, Romeijn JW. Psychiatric comorbidity: fact or artifact? *Theor Med Bioeth.* (2015) 36:41–60. doi: 10.1007/s11017-015-9321-0
48. Cloninger CR, Cloninger CR. A unified biosocial theory of personality and its role in the development of anxiety states: a reply to commentaries. *Psychiatr Dev.* (1988) 6:83–120.
49. Amorim P, Lecrubier Y, Weiller E, Hergueta T, Sheehan D. DSM-III-R Psychotic Disorders: procedural validity of the Mini International Neuropsychiatric Interview (MINI). Concordance and causes for discordance with the CID. *Eur Psychiatry.* (1998) 13:26–34. doi: 10.1016/S0924-9338(97)86748-X
50. Beck AT, Steer RA, Ball R, Ranieri WF. Comparison of beck depression inventories-IA and-II in psychiatric outpatients. *J Pers Assess.* (1996) 67:588–97. doi: 10.1207/s15327752jpa6703_13
51. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. *J Consult Clin Psychol.* (1988) 56:893–7. doi: 10.1037//0022-006x.56.6.893
52. Harman HH. *Modern Factor Analysis*. Chicago, IL, USA: University of Chicago Press (1976).
53. Rosseel Y. lavaan: an R package for structural equation modeling. *J Stat Softw.* (2012) 48:1–36. doi: 10.18637/JSS.V048.I02
54. Melegari MG, Innocenzi M, Marano A, De Rosa L, Donfrancesco R, Rozza S, et al. Application of the seven-factor-model of personality to an Italian preschool sample. *Psychiatry Investig.* (2014) 11:419–29. doi: 10.4306/pi.2014.11.4.419
55. Sahonero-Alvarez G, Calderon H. A comparison of SOBI, fastICA, JADE and infomax algorithms, in: *Proceedings of The 8th International Multi-Conference on Complexity, Informatics and Cybernetics*. (2017) pp. 17–22. Available online at: <https://www.iis.org/CDs2017/CD2017Spring/papers/ZA832BA.pdf>.
56. Boroomand A, Ahmadian A, Oghabian MA, Alirezaie J, Beckman C. An efficient hybrid wavelet-ICA algorithm for analyzing simulated fMRI data in noisy environment, in: *2007 IEEE International Symposium on Signal Processing and Information Technology*, Giza, Egypt. (2007) pp. 408–13. doi: 10.1109/ISSPIT.2007.4458163
57. Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika.* (1965) 30:179–85. doi: 10.1007/bf02289447
58. Revelle W. *psych: Procedures for Psychological, Psychometric, and Personality Research* (2022). Available online at: <https://CRAN.R-project.org/package=psych> (accessed May 20, 2024).
59. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Of Artif Intell Res.* (2011) 16:321–57. doi: 10.1613/jair.953
60. Sui Y, Wei Y, Zhao D. Computer-aided lung nodule recognition by SVM classifier based on combination of random undersampling and SMOTE. *Comput Math Methods Med.* (2015) 2015:1–13. doi: 10.1155/2015/368674
61. Lemaitre G, Nogueira F, Aridas CK. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *J Mach Learn Res.* (2017) 18:1–5. <http://jmlr.org/papers/v18/16-365.html>.
62. Noble WS. What is a support vector machine? *Nat Biotechnol.* (2006) 24:12. doi: 10.1038/nbt1206-1565
63. Lameski P, Zdravetski E, Mingov R, Kulakov A. SVM parameter tuning with grid search and its impact on reduction of model over-fitting. *Lecture Notes Comput Sci (including subseries Lecture Notes Artif Intell Lecture Notes Bioinformatics)*. (2015) 9437:464–74. doi: 10.1007/978-3-319-25783-9_41
64. De Raad B, Mlačić B. The lexical foundation of the Big Five factor model. In: Widiger TA, editor. *The Oxford handbook of the five factor model*. New York: Oxford University Press (2017). p. 191–216.

65. Trofimova I, Robbins TW, Sulis WH, Uher J. Taxonomies of psychological individual differences: Biological perspectives on millennia-long challenges. *Philos Trans R Soc B: Biol Sci.* (2018) 373. doi: 10.1098/rstb.2017.0152
66. Rusalov VM, Trofimova IN. *Structure of temperament and its measurement*. Toronto, Canada: PSP: Psychological Services Press (2007).
67. Trofimova IN, Sulis W. A study of the coupling of FET temperament traits with major depression. *Front Psychol.* (2016) 7:1848. doi: 10.3389/fpsyg.2016.01848
68. Widiger TA, McCabe GA. The alternative model of personality disorders (AMPD) from the perspective of the five-factor model. *Psychopathology.* (2020) 53:149–56. doi: 10.1159/000507378
69. Earleywine M. The examination of alternative models enhances confirmatory factor analyses. *Pers Individ Dif.* (1993) 15:593–4. doi: 10.1016/0191-8869(93)90343-2
70. Hyvarinen A, Karhunen J, Oja E. What is Independent Component Analysis? In: Hyvarinen A, Karhunen J, Oja E, editors. *Independent Component Analysis*. New York, NY, USA: Wiley-Interscience, John Wiley & Sons, Inc. (2001). p. 147–64.
71. Craske MG, Herzallah MM, Nusslock R, Patel V. From neural circuits to communities: an integrative multidisciplinary roadmap for global mental health. *Nat Ment Health.* (2023) 1:1. doi: 10.1038/s44220-022-00012-w
72. Sulis W. Quo vadis taxonomies of consistent behavioural patterns: time to change horses? *Curr Opin Behav Sci.* (2022) 44. doi: 10.1016/j.cobeha.2022.101103