Journal of Psychiatric Research 141 (2021) 199–205

Contents lists available at ScienceDirect

Journal of Psychiatric Research

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

Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders

Thalia Richter [a,*], Barak Fishbain [b], Eyal Fruchter [c], Gal Richter-Levin [a], Hadas Okon-Singer [a]

[a] Department of Psychology, School of Psychological Sciences, University of Haifa, Mount Carmel Haifa, Israel
[b] Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa, Israel
[c] Brus Rappaport Faculty of Medicine, Technion– Israel Institute of Technology, Haifa, Israel

A R T I C L E I N F O

Keywords:
Machine learning
Anxiety
Depression
Psychiatric diagnosis
Cognitive bias
Diagnostic support system

A B S T R A C T

In light of the need for objective mechanism-based diagnostic tools, the current research describes a novel diagnostic support system aimed to differentiate between anxiety and depression disorders in a clinical sample.

Eighty-six psychiatric patients with clinical anxiety and/or depression were recruited from a public hospital and assigned to one of the experimental groups: Depression, Anxiety, or Mixed. The control group included 25 participants with no psychiatric diagnosis.

Participants performed a battery of six cognitive-behavioral tasks assessing biases of attention, expectancies, memory, interpretation and executive functions. Data were analyzed with a machine-learning (ML) random forest-based algorithm and cross-validation techniques. The model assigned participants to clinical groups based solely on their aggregated cognitive performance.

By detecting each group's unique performance pattern and the specific measures contributing to the prediction, the ML algorithm predicted diagnosis classification in two models: (I) anxiety/depression/mixed vs. control (76.81% specificity, 69.66% sensitivity), and (II) anxiety group vs. depression group (80.50% and 66.46% success rates in classifying anxiety and depression, respectively).

The findings demonstrate that the cognitive battery can be utilized as a support system for psychiatric diagnosis alongside the clinical interview. This implicit tool, which is not based on self-report, is expected to enable the clinician to achieve increased diagnostic specificity and precision. Further, this tool may increase the confidence of both clinician and patient in the diagnosis by equipping them with an objective assessment tool. Finally, the battery provides a profile of biased cognitions that characterizes the patient, which in turn enables more fine-tuned, individually-tailored therapy.

1. Introduction

In recent years, various research efforts have been invested in developing novel and objective methods for psychiatric diagnosis based on disorder-related mechanisms, rather than on the self-reported symptoms-based diagnosis commonly used today (Hofmann and Hayes 2020). An example of one such effort is the development of the NIMH Research Domain Criteria (RDoC), in which classification is done based on behavioral dimensions and neurobiological measures (Cuthbert and Insel, 2013). The need for new classification methods for mental disorders is highly salient in the case of anxiety and depressive mood disorders, due to their high prevalence in the general population (Wu and Fang, 2014), the high comorbidity rates between them (Van den Bergh et al., 2020) and the average to low treatment success rates over time, manifested by patients' relapses and chronic episodes (Hunsley et al., 2013). The notion of diagnosing through detection of specific disorder-related mechanisms, such as cognitive biases presented by individuals suffering from psychopathology, emerged from substantial findings regarding their prominent role in the onset, maintenance and possibility of recovery from these conditions, established in both past and current research (Beck, 1967; Power and Dalgleish, 2015).

Recently, we demonstrated success in differentiation between anxiety and depression, using a new diagnostic tool (Richter et al., 2020), developed by us and validated in a nonclinical study. The tool comprises six cognitive-behavioral tasks, each examining a different type of bias (selective attention, spatial attention, expectancy, interpretation,

* Corresponding author. Department of Psychology, University of Haifa, Mount Carmel Haifa, 3498838, Israel.
E-mail address: thalia.richter173@gmail.com (T. Richter).

https://doi.org/10.1016/j.jpsychires.2021.06.044
Received 28 January 2021; Received in revised form 20 May 2021; Accepted 22 June 2021
Available online 24 June 2021

memory and cognitive control deficiency). The novelty of this tool lies in its combination of various tasks, and their analysis using a designated machine-learning (ML) algorithm, which enables prediction of the probability of diagnosis of anxiety/depression/healthy. The research was performed among subclinical participants, assigned by questionnaires to one of four groups – high level of anxiety symptoms, high level of depression symptoms, high levels of both anxiety and depression symptoms, and controls with low levels of anxiety and depression symptoms. The algorithm predicts each participant's group affiliation based on their pattern of behavioral performance in the tasks, and its relative resemblance to the characteristic performance pattern for each group. Further, the algorithm reports the contribution of each measure, obtained via the test battery, to the prediction. This information allows better understanding of the manifestation of cognitive biases among depressed and anxious individuals, and elucidates the specific biases that differ most and least between groups. To summarize, this new methodological paradigm has yielded an algorithm that is promising and performs well. We now seek to apply the same methodological paradigm to investigate whether it yields an algorithm that performs equally well in a clinical sample.

Thus, the aim of the current study is to replicate the findings in a clinical population, a step that is essential for determining the ability of the paradigm to predict a psychiatric diagnosis. Further, the study aims to establish the status of the paradigm as a diagnostic tool, facilitating its assimilation in the psychiatric diagnosis process in the future. The possible benefits of utilizing this battery alongside the clinical interview may be increased diagnostic specificity and precision, leading to a more fine-tuned, individually-tailored therapy. Further, it may assist in increasing clinicians' confidence in their diagnoses, by equipping them with a more objective tool of assessment.

ML analysis has been gaining popularity in mental health research in recent years. For example, in a recent study (Chekroud et al., 2016), implementation of the variables that were found by ML algorithms to increase treatment efficacy indeed resulted in improved treatment for depression. The current study sought to examine the possibility of differentiating between anxiety and depression by detecting a unique pattern of biased reactions to emotional stimuli that characterize each disorder. Several methodologies can be utilized for this purpose, among them multivariate and logistic regressions as well as power analysis using ANOVA (e.g., see Baxter et al., 2013; Beiter et al., 2015; Cape et al., 2010). However, these methods all assume an underlying mathematical model that does not necessarily comply with human behavior. Both neural networks (Kumar et al., 2020) and regression trees (Kessler et al., 2016) eliminate the need to discern underlying models. Neural networks often entail a heavy computational undertake that hinders their use. The use of a random forest facilitates accurate classification in the presence of large variance in the measures within each group because the forest allows for a large number of repetitions (trees) in the classification process (Kanchanatawan et al., 2018). In this way, even subtle trends manifest themselves in the classification results (Kessler et al., 2016). Further, recent studies that aimed to diagnose depression and anxiety examined the efficacy of several ML methods. Random forest was found repeatedly as the most accurate model (de Souza Filho et al., 2021; Priya et al., 2020). Thus, the random forest algorithm is most suitable for the task at hand.

To the best of our knowledge, the current study is the first project to differentiate anxiety and depression through the detection of cognitive-behavioral patterns by ML procedures. As self-reports tend to be biased by individuals' own perception of their condition (Bauhoff, 2011), the possibility of adding this tool, which is completely data-driven, to the diagnostic process may add further objectivity in the differential diagnosis decision, thus leading to better personalized treatment.

2. Method

2.1. Participants

Participants in the study were 111 individuals who use Hebrew as their main language and have normal or corrected-to-normal vision. Ten participants did not finish the study, resulting in a total sample size of 101 participants. The majority of participants were women, similarly to the proportion of diagnosis of women versus man in anxiety and depression disorders (Faravelli et al., 2013). The study was approved by the Helsinki committee of Rambam Health Care Campus (0440-17-RMB). Participants signed an informed consent form prior to participation and were debriefed at the end of the experiment. Participants received monetary compensation for their time.

Table 1 provides demographic and clinical information by group.

Clinical patients were recruited from the mental health department of Rambam Health Care Campus in Haifa. Additional clinical patients as well as the control group were recruited through advertisements posted around the hospital, and uploaded to relevant social media groups, inviting individuals with or without clinical diagnosis to take part. Control participants reported they have no current or past psychiatric history. Clinical patients were initially interviewed by a clinically-trained psychologist using the Structured Clinical Interview for DSM-5 (SCID-5; First et al., 2015), to confirm diagnosis, and were accordingly assigned to one of the following experimental groups: Depression (MDD, dysphoria, adjustment disorder with depressive reaction, depressive episode); Anxiety (specific phobia, social anxiety, panic attacks {with or without agoraphobia}), GAD, Anxiety-NOS, adjustment disorder with anxious reaction); Mixed; or Control. To control for factors that might interfere with the results, participants were asked about their neurological and psychiatric history (epilepsy or hemophilia conditions, past and current depression/anxiety diagnoses and related prescribed

Table 1
Mean and standard deviations (in parenthesis) of demographic and clinical information by group.

Group	Anxiety (n = 24)	Depression (n = 25)	Mixed (n = 27)	Control (n = 25)
Age	33 (9.86)	47 (15.87)	32 (11.50)	27 (5.96)
Anxiety		$p < .05$	n.s.	n.s.
Depression	$p < .05$		$p < .05$	$p < .05$
Mixed	n.s.	$p < .05$		n.s
Control	n.s.	$p < .05$	n.s	
Sex	19 women (79%)	17 women (68%)	18 women (66%)	16 women (64%)
Education (years)	13 (2.30)	13 (2.22)	13 (2.51)	14 (2.08)
Anxiety		n.s.	n.s.	n.s.
Depression	n.s.		n.s.	n.s.
Mixed	n.s.	n.s.		n.s.
Control	n.s.	n.s.	n.s.	
BDI score	21 (12.36)	27 (13.42)	29 (12.13)	7 (7.61)
Anxiety		n.s.	n.s.	$p < .05$
Depression	n.s.		n.s.	$p < .05$
Mixed	n.s.	n.s.		$p < .05$
Control	$p < .05$	$p < .05$	$p < .05$	
STAI score	54 (9.90)	53 (9.95)	57 (10.41)	39 (9.76)
Anxiety		n.s.	n.s.	$p < .05$
Depression	n.s.		n.s.	$p < .05$
Mixed	n.s.	n.s.		$p < .05$
Control	$p < .05$	$p < .05$	$p < .05$	
PSWQ score	62 (11.74)	56 (9.59)	60 (12.98)	46 (13.84)
Anxiety		n.s.	n.s.	$p < .05$
Depression	n.s.		n.s.	$p < .05$
Mixed	n.s.	n.s.		$p < .05$
Control	$p < .05$	$p < .05$	$p < .05$	
RRS score	54 (14.04)	58 (15.59)	62 (11.83)	38 (12.51)
Anxiety		n.s.	n.s.	$p < .05$
Depression	n.s.		n.s.	$p < .05$
Mixed	n.s.	n.s.		$p < .05$
Control	$p < .05$	$p < .05$	$p < .05$	

medications, or any other psychiatric or neurological history, syndromes or diseases), as well as the presence of any learning disabilities or attention deficit disorders.

2.2. Questionnaires

Levels of anxiety symptoms, depression symptoms, rumination and worry were assessed by the following questionnaires, respectively: the State–Trait Anxiety Inventory–Trait Version (STAI-T; Spielberger et al., 1983), the Beck Depression Inventory–Second Edition (BDI–II; Beck and Beamesderfer, 1974), the Ruminative Responses Scale (RRS; Nolen--Hoeksema and Morrow, 1991), and the Penn State Worry Questionnaire (PSWQ; Meyer et al., 1990).

2.3. Behavioral measurements

Readers are referred to the supplementary material section for further information about each task.

Cognitive biases were measured using a test battery of cognitive-behavioral tasks that was developed by our group and previously validated among individuals with subclinical levels of depression and anxiety symptoms (Richter et al., 2020). As noted, in the current study the same battery was applied again to test performance among a clinical population. The battery comprises six tasks. Each task measures a specific cognitive bias using a prevalent paradigm, with modifications allowing it to test for both automatic and non-automatic reactions. Reaction times (RTs), accuracy rates and explicit selection reactions were recorded:

Biases in selective attention (the ability to ignore distracting emotional information) – measured by the focused attention flanker task (FAFT; Lichtenstein-Vidne et al., 2012).
Biases in spatial attention (abnormal orienting of attention to negative stimuli, either enhanced orienting or avoidance) – measured by the emotional dot-probe task (EDPT; MacLeod et al., 1986).
Expectancy biases (abnormal expectation of negative future events) – measured by the future events task (FET; Miranda and Mennin, 2007).
Biases in interpretation (interpretation by anxious or depressed participants compared to healthy participants of ambiguous situations as more negative) – measured by the word–sentence association paradigm (WSAP; Beard and Amir, 2009).
Biases in memory (abnormal memory of emotional versus neutral items, compared to healthy participants) – measured by the word identification task (WIT; Tarsia et al., 2003).
Cognitive control – measured by the internal switching task (IST; Beckwé et al., 2014).

Participants took part in the study during two sessions of approximately 1 h each, with an interval of up to two weeks between sessions.

2.4. Data analysis

Participants data goes under data cleansing and ML analysis. ML is a well-established branch of computer analysis techniques that are used to classify large numbers of observation-based patterns in the input data. Supervised machines (i.e., algorithms) are trained on a set of observations belonging to known classes; i.e., the training set. In this case random 80% of the participants were selected to be of the training set. The output of this training process is a classifier. The classifier is a set of rules by which unclassified, new, measurement is categorized. Random forest is a supervised multi-class classifier based on a collection of decision trees (Breiman, 2001; Breiman et al., 1984) The random forest classifier (RFC) uses voting between an ensemble of decision trees (hence "random forest"). During the training stage of RFC, a random subset of the training set observations is chosen and is left unused (out-of-bag observations). The out-of-bag-observations are then employed during the derivation of the classifier to assess its performance and the effect of parameters, such as the number of trees, on the classification error. The reminder of the patients (20%) were used here as out-of-the-bag observations. In order to avoid any bias due to a specific selection of the training set, the training-validation process is repeated 1000 times. Each time with different randomly selected training set. To account for different group sizes when needed, the number of participants selected from each group is equalized to the smallest group. This process of selection is done 10 times. In each selection, randomly different participants are chosen from each group according to the number set by the smallest one. classification results were assessed by Confusion matrices (Stehman, 1997) and McNemar's test (Stuart, 1955). The contribution of each feature (e.g. its importance) to classification accuracy was evaluated (Friedman, 2001; Richter et al., 2020). Feature importance is a valuable tool for designing future studies, for simplifying the measurement system, by setting it to measure solely the features which were deemed important. Further, the features that were found to be potentially important can shed light on the phenomena under investigation (Huynh-Thu et al., 2012). The algorithm was implemented in Matlab (2020). Readers are kindly referred to the supplementary material for information about the data cleaning process, missing data, and elaborated description of the ML algorithm.

3. Results

3.1. Comparing patients to healthy controls (depression + anxiety + mixed groups versus control group model)

The leave-one-out analysis (omitting one measure each time) revealed the marginal contribution of each behavioral measure drawn from the input. Fig. 1 shows the normalized error differences between the classification of all the behavioral measures, with and without the specific measure. The larger the difference is, the larger the unique contribution of the specific behavioral measure.

The *bagged decision tree classification* algorithm demonstrated 76.81% prediction success for the healthy control group, and 69.66% prediction success for participants in the clinical patients' groups. Table 2 shows the classification accuracy of the participants in each cluster.

McNemar's test, aimed at evaluating paired dichotomous data (Adedokun and Burgess, 2012) demonstrated the classifier's higher distinguishing between group ability compared to random classification ($\chi2(1) = 4.91$, $p = .03390033$).

3.2. Comparing the anxiety versus depression model

To test the strength of the algorithm in differentiating solely between depression vs. anxiety groups, a second two-group model was introduced into the classification scheme.

The *leave-one-out* analysis revealed the marginal contribution of each behavioral measure drawn from the input, as shown in Fig. 2.

The *bagged decision tree classification* algorithm revealed 80.50% success in classifying participants in the anxiety group, and 66.46% success in classifying participants in the depression group. Table 3 shows the accuracy of the classification of participants into each group.

McNemar's test shows the classifier's higher ability to distinguish between the groups compared to random classification ($\chi2(1) = 6.25$, $p = .00788231$).

3.3. Examining the uniqueness of the mixed group

Two separate analyses were conducted to test the extent to which the mixed group is distinguished from each other diagnosis.

The *bagged decision tree classification* algorithm revealed 64.39% and 73.66% success in classifying participants in the mixed group as opposed to the anxiety and depression groups, respectively. Classification rates of

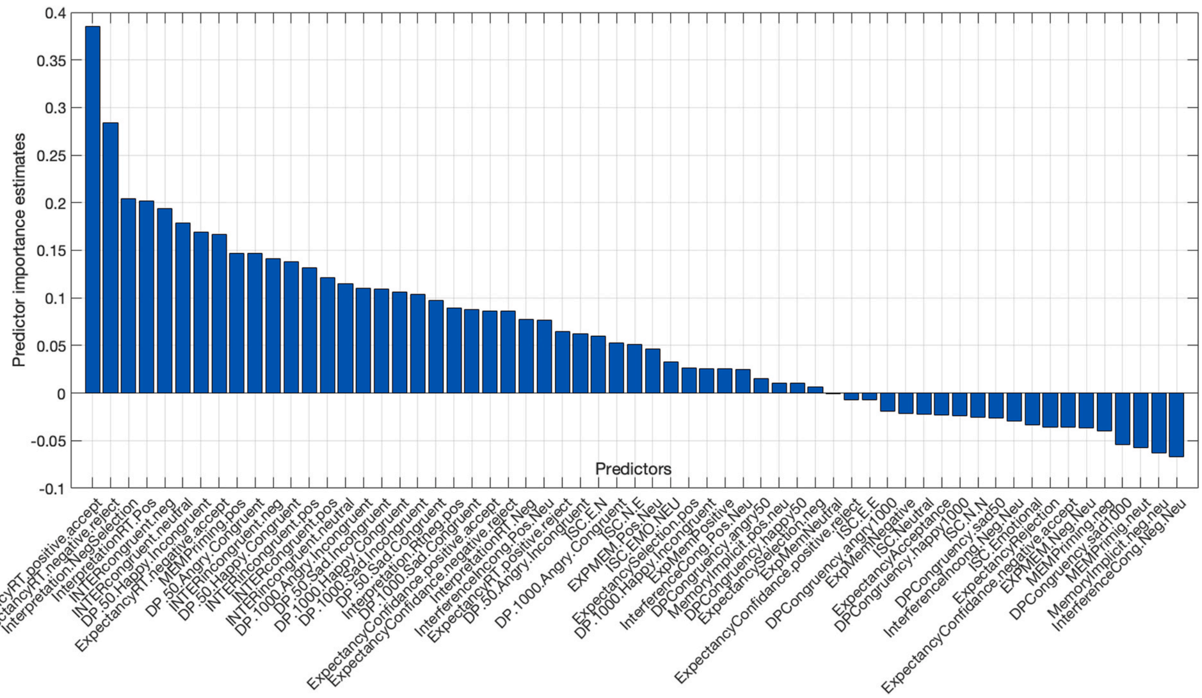

Fig. 1. Marginal contribution of each behavioral measure. The normalized error difference between the classification of all the behavioral features, with and without each specific measure as indicated on the X axis, is shown. The larger the difference is, the larger the unique contribution of the specific behavioral measure. Error bars represent the standard deviations of results between iterations. Measures that are located to the left of the zero-point showed a positive contribution to the model and were inserted into the bagged tree algorithm. Thus, after performing the leave-one-out analysis, only measures that made a positive contribution were used. Out of 61 behavioral measures that were analyzed, 39 were found to contribute to the model prediction (22 did not). As seen on the left of the X axis, the measure contributing the most to the classification was the RT of acceptance of positive situations in the FET (which measures expectancy biases), followed by the RT of rejection of negative situations in the same task. The least contributing measure was a FAFT (which measures selective attention biases) index for the effect of negative distractors in congruent trials.

Table 2
Confusion matrix showing the classification accuracy. Accuracy rates represent the classification prediction of each participant, which was based on the participant's behavioral measures, as compared to their true group membership, which was based on their diagnosis. Each matrix row represents the instances in a predicted class while each column represents the instances in an actual class. The correctly classified members of the dataset reside along the matrix's diagonal. The algorithm's accuracy is then measured by the trace of the matrix divided by its sum of all elements.

Predicted	Anxiety + Depression + Mixed	Control
True		
Anxiety + Depression + Mixed	**0.6966**	0.3034
Control	0.2319	**0.7681**

the other groups versus the mixed group were 64.11% for anxiety and 71.35% for depression.

Table 4 shows the accuracy of classifying the participants into each group compared the mixed group. The bottom row represents the success rate in correctly classifying mixed participants versus the anxiety and the depression group, respectively.

McNemar's test shows that the classifier is better able to distinguish between the depression and the mixed group than random classification- $\chi2(1) = 6.74$, $p = .001163242$. However, using McNemar's test to distinguish between the anxiety and the mixed group exhibits no difference from random classification- $\chi2(1) = 2.88$, p = n.s).

4. Discussion

The current study aimed to answer the need for novel and objective psychiatric diagnosis tools, by differentiating between clinical patients with anxiety and/or depression disorders, using cognitive-behavioral performance data and advanced ML analysis tools. This combination enabled the detection of a unique pattern of biased reactions to emotional stimuli in each disorder, based on participants' aggregated performance in the behavioral tasks. Selective and spatial attention, expectancy, interpretation and memory biases were examined, as well as cognitive control deficiencies. The analysis reached 76.81% prediction success in the healthy control group, and 69.66% prediction success for participants in the clinical patients' groups. Analysis focusing on anxiety vs. depression yielded 80.50% success in classifying participants in the anxiety group, and 66.46% success in classifying the depression group. These classification accuracies were all above chance level.

Another aim of this study was to replicate the paradigm's previous findings among subclinical anxious and depressed individuals and to test its ability to perform equally well in a clinical sample. The current results are in line with our previous study, presenting similar success levels in both analysis models (Richter et al., 2020). The repeated success at differentiating between anxiety and depression based solely on behavioral reactions, without any self-report measures, points to the paradigm's consistency and stability. This is also in accordance with previous studies suggesting that bias effects exist in both clinical and subclinical populations (Bar-Haim et al., 2007; Gaddy and Ingram, 2014; Kircanski and Gotlib, 2015).

As can be seen in Figs. 1 and 2, and in accordance with Richter et al. (2020) measures from all the tasks contributed to the classification prediction in both models. This finding provides further evidence of the existence of various cognitive biases among individuals with anxiety and depression disorders. Moreover, the finding underscores the need to investigate the field of cognitive biases among psychopathological populations through a broad perspective, considering the combination of performance patterns in different functions. The benefits of ML

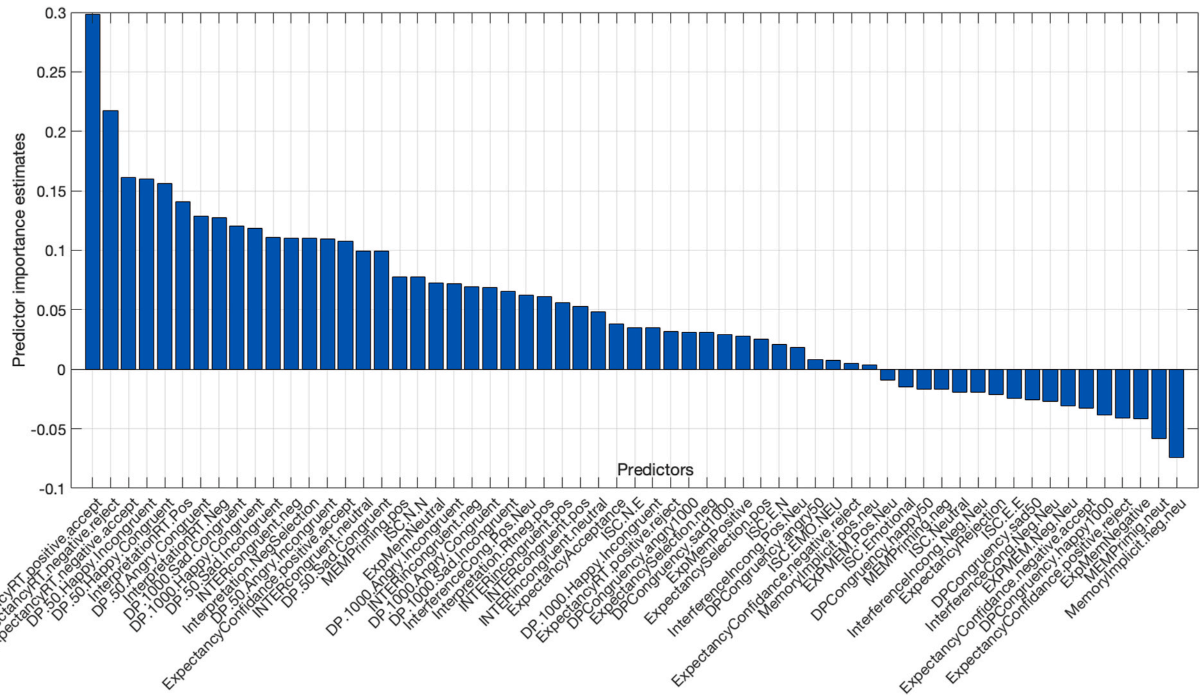

Fig. 2. Marginal contribution of each behavioral measure. The normalized error difference between the classification of all the behavioral features, with and without the specific measure on the X axis, is shown. The larger the difference is, the larger the unique contribution. Measures that are located to the left of the zero-point demonstrated a positive contribution to the model and were inserted into the bagged tree algorithm. Out of 61 behavioral measures that were analyzed, 41 were found to contribute to the prediction success, while 20 were not. For example, the most-contributing measure was the RT of acceptance of positive situations in the FET (which measures expectancy biases), followed by the RT of rejection of negative situations in the same task. The least-contributing measure that was not inserted into the algorithm was a WIT (which measures memory biases) index of the relative implicit remembering of negative words over neutral words.

Table 3
Accuracy rates represent the classification prediction of each participant, which was based on their individual behavioral measures, as compared to the true group membership, which was based on their diagnosis.

Predicted / True	Anxiety	Depression
Anxiety	**0.8050**	0.1950
Depression	0.3354	**0.6646**

Table 4
Accuracy rates represent the classification prediction of each participant based on individual behavioral measures, as compared to true group membership, which is based on diagnosis.

Predicted / True	Anxiety	Depression	Mixed
Anxiety	**0.6411**		0.3589
Depression		**0.7135**	0.2865
Mixed	0.3561	0.2634	
Mixed (correct classification)	**0.6439**	**0.7366**	

analyses lie in its ability to connect and compare data from various bias categories, to create a united pattern characterizing each group.

Our results advance the possibility of the battery being utilized during psychiatric diagnosis, adding higher confidence to the clinician's decision that can now be based on a tool less prone to self-report biases. The need for more objective tools in the diagnostic procedure was discussed by Cuthbert and Insel (2013). Their research stated that this shortcoming results in clinical implications that include steady prevalence and mortality rates for any mental illness over time, as opposed to physical illnesses that often show decreasing prevalence and mortality rates. Another implication is the observed low effectiveness of psychotherapy treatments and medications, such as SSRIs, as they are given to a broad spectrum of patients due to "artificial grouping of heterogeneous syndromes with different pathophysiological mechanisms into one disorder". Thus, a more objective tool of assessment, which helps reach the correct differential diagnosis, may contribute to the development of more accurate and individual-specific treatments.

Moreover, the repeated success in distinguishing between anxiety and depression, based solely on cognitive biases measures, corresponds to recent findings regarding neurocognitive processes as mediators of psychological treatment effects (Reiter et al., pre-print 2020). In their meta-analysis, however, Hallion and Ruscio (2011) pointed to the decreasing efficacy of cognition-based treatment, specifically cognitive-behavioral modification (CBM), when examining its cumulative effects on anxiety and depression, separately. This may be due to paradigms not designated to be disorder-specific to begin with, tapping more of the shared biases of anxiety and depression instead of the difference in the biases. Future research may benefit from the current study because of its disease-specific feature selection.

The ML algorithm demonstrated greater success at predicting anxiety than depression. This may be due to the possibility of depression sometimes being secondary to severe anxiety (Meier et al., 2015) (i.e., developing depression as a result of being unable to cope with anxiety without treatment). Secondary depression will be manifested in depressive symptoms and may, therefore, be diagnosed as depression. Nonetheless, the underlying cognitive mechanisms and resulting performance patterns may be more similar to those related to anxiety, and, therefore, these participants may be "mistakenly" recognized as anxious by the ML algorithm.

This conclusion gains further support from the analyses that tested the uniqueness of the mixed anxiety-depression group. The results showed that the performance patterns of the depression group differ significantly from those of the mixed group, while the performance of the anxiety group did not differ significantly from that of the mixed group. Classification rates for depressed [71.35%] versus mixed [73.66%] participants were similar to those for anxious versus depressed participants. These results corroborate the notion that a mixed anxiety-depression diagnosis can often result from worsening anxiety and therefore can exhibit similar performance patterns, while depression may be characterized by different impairments of cognitive mechanisms, resulting in a distinguished pattern of performance.

The main limitation of the current study is the rather small sample, when considering each group separately. This gains further importance as the findings were not validated by an external dataset, but by cross validation procedures. However, as mentioned the diagnostic battery was already validated in a sub-clinical population with similar success rates (Richter et al., 2020). This strengthen the consistency of the algorithm and reduces the possibility of over-fitting. Future studies are planned in order to validate the findings in external datasets.

Similar to other studies using ML classification, another limitation lies in the fact that the cognitive battery is validated using a psychiatric diagnosis that is inherently prone to self-report biases. This bias, however, is mitigated by the fact that the participant's classification is not evaluated solely on the basis of their diagnosis, but also on how it compares to all the diagnoses of all the other participants. In the future, studies may overcome this inherent limitation by creating a constantly updating database, uniting all input received from studies that will utilize the paradigm, in order for the algorithm to keep re-validating itself. This database will allow the continual updating and refinement of the characteristic patterns of each disorder and the predicted diagnosis, as currently done, for example, in one the most prevalent personality-diagnosis questionnaires, the Minnesota Multiphasic Personality Inventory (MMPI; Butcher, 2011).

Further, significant age differences were found between the study groups (see Table 1). Therefore, another ML analysis was conducted for both models with age as one of the features in order to test whether the effect of age may underlie the differences discovered between performance patterns. Although age was found to contribute to the prediction by a 'leave-one-out' analysis, it almost did not change the prediction accuracies and even slightly reduced them. Further, no significant age differences between the groups emerged in our previous study (Richter et al., 2020) and the prediction accuracies were rather similar to those in the current study. Therefore, it is reasonable to assume that any age effect on prediction is specific to the current sample and cannot be generalized. Future studies should focus on exploring the effects of age by classifying participants into different age groups.

Finally, because most of the participants were female, the effect of sex could not be estimated in the study. Future research is encouraged to examine the implications of sex.

In sum, the current study reinforced the validity of the combination of the cognitive-behavioral test battery analyzed by the ML algorithm, as a diagnostic tool, capable of differentiating between anxiety and depression, based on cognitive-behavioral performance patterns and no self-report measures, in order to contribute to differential diagnostic decisions. Because this is the first time this tool was applied on a clinical population, further research is needed in order to test its generalizability and determine its promising status as a diagnostic support system for clinicians, leading towards more specific and refined diagnoses and clinical treatments.

CRediT authorship contribution statement

Thalia Richter: Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing. **Barak Fishbain:** Software, Formal analysis, Data curation, Writing – review & editing. **Eyal Fruchter:** Resources, Writing – review & editing, Supervision, Project administration. **Gal Richter-Levin:** Conceptualization, Writing – review & editing, Resources. **Hadas Okon-Singer:** Conceptualization, Methodology, Writing – review & editing, Visualization, Supervision, Funding acquisition.

Declaration of competing interest

Authors declare this research was not supported by any funding or sponsorship, and that no conflict of interests exists.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpsychires.2021.06.044.

References

Adedokun, O.A., Burgess, W.D., 2012. Analysis of paired dichotomous data: a gentle introduction to the McNemar test in SPSS. J. Multidiscip. Res. 8, 125–131.
Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M.J., Van Ijzendoorn, M. H., 2007. Threat-related attentional bias in anxious and non- anxious individuals: a meta-analytic study. Psychol. Bull. 133, 1–24. https://doi.org/10.1037/0033-2909.133.1.1.
Bauhoff, S., 2011. Systematic self-report bias in health data: impact on estimating cross-sectional and treatment effects. Health Serv. Outcome Res. Methodol. 11, 44–53. https://doi.org/10.1007/s10742-011-0069-3.
Baxter, A.J., Scott, K.M., Vos, T., Whiteford, H.A., 2013. Global prevalence of anxiety disorders: a systematic review and meta-regression. Psychol. Med. 43 (5), 897. https://doi.org/10.1017/S003329171200147X.
Beard, C., Amir, N., 2009. Interpretation in social anxiety: when meaning precedes ambiguity. Cognit. Ther. Res. 33, 406–415. https://doi.org/10.1007/s10608-009-9235-0.
Beck, A.T., 1967. Depression: Clinical, Experimental, and Theoretical Aspects. University of Pennsylvania Press.
Beck, A.T., Beamesderfer, A., 1974. Assessment of depression: the depression inventory. In: Pichot, P., Olivier-Martin, R. (Eds.), Psychological Measurements in Psychopharmacology. Karger.
Beckwé, M., Deroost, N., Koster, E.H., De Lissnyder, E., De Raedt, R., 2014. Worrying and rumination are both associated with reduced cognitive control. Psychol Res. 78, 651–660. https://doi.org/10.1007/s00426-013-0517-5.
Beiter, R., Nash, R., McCrady, M., Rhoades, D., Linscomb, M., Clarahan, M., Sammut, S., 2015. The prevalence and correlates of depression, anxiety, and stress in a sample of college students. J. Affect. Disord. 173, 90–96. https://doi.org/10.1016/j.jad.2014.10.054.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Routledge.
Butcher, J.N., 2011. A Beginner's Guide to the MMPI-2. American Psychological Association.
Cape, J., Whittington, C., Buszewicz, M., Wallace, P., Underwood, L., 2010. Brief psychological therapies for anxiety and depression in primary care: meta- analysis and meta-regression. BMC Med. 8 (1), 1–13. https://doi.org/10.1186/1741-7015-8-38.
Chekroud, A.M., Zotti, R.J., Shehzad, Z., et al., 2016. Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry 3, 243–250. https://doi.org/10.1016/S2215-0366(15)00471-X.
Cuthbert, B.N., Insel, T.R., 2013. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC Med. 11, 126. https://doi.org/10.1186/1741- 7015-11-126.
de Souza Filho, E.M., Rey, H.C.V., Frajtag, R.M., Cook, D.M.A., de Carvalho, L.N.D., Ribeiro, A.L.P., Amaral, J., 2021. Can machine learning be useful as a screening tool for depression in primary care? J. Psychiatr. Res. 132, 1–6.
Faravelli, C., Scarpato, M.A., Castellini, G., Sauro, C.L., 2013. Gender differences in depression and anxiety: the role of age. Psychiatr. Res. 210, 1301–1303.
First, M.B., Williams, J.B., Karg, R.S., Spitzer, R.L., 2015. Structured Clinical Interview for DSM-5—Research Version (SCID-5 for DSM-5, Research Version; SCID-5-RV). American Psychiatric Association, Arlington, VA.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 129, 1189–1232. https://doi.org/10.1214/aos/1013203451.
Gaddy, M.A., Ingram, R.E., 2014. A meta-analytic review of mood-congruent implicit memory in depressed mood. Clin. Psychol. Rev. 34, 402–416. https://doi.org/10.1016/j.cpr.2014.06.001.
Hallion, L.S., Ruscio, A.M., 2011. A meta-analysis of the effect of cognitive bias modification on anxiety and depression. Psychol. Bull. 137, 940–958. https://doi.org/10.1037/a0024355.
Hofmann, S.G., Hayes, S.C., 2020. Beyond the DSM: toward a Process-Based Alternative for Diagnosis and Mental Health Treatment. Context Press/New Harbinger Publications, Oakland, CA.
Hunsley, J., Elliott, K., Therrien, Z., 2013. The Efficacy and Effectiveness of Psychological Treatments. Canadian Psychological Association, Ottawa, Canada.
Huynh-Thu, V.A., Saeys, Y., Wehenkel, L., Geurts, P., 2012. Statistical interpretation of machine learning-based feature importance scores for biomarker discovery. Bioinformatics 28, 1766–1774. https://doi.org/10.1093/bioinformatics/bts238.
Kanchanatawan, B., Thika, S., Sirivichayakul, S., Carvalho, A.F., Geffard, M., Maes, M., 2018. schizophrenia, depression, anxiety, and physiosomatic symptoms are strongly

related to psychotic symptoms and excitation, impairments in episodic memory, and increased production of neurotoxic tryptophan catabolites: a multivariate and machine learning study. Neurotox. Res. 33, 641–655. https://doi.org/10.1007/s12640-018-9868-4.
Kessler, R.C., van Loo, H.M., Wardenaar, K.J., Bossarte, R.M., Brenner, L.A., Cai, T., et al., 2016. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. Mol. Psychiatr. 21 (10), 1366–1371. https://doi.org/10.1038/mp.2015.198.
Kircanski, K., Gotlib, I.H., 2015. Processing of emotional information in major depressive disorder: toward a dimensional understanding. Emot Rev 7, 256–264. https://doi.org/10.1177/1754073915575402.
Kumar, P., Garg, S., Garg, A., 2020. Assessment of anxiety, depression and stress using machine learning models. Procedia Computer Science 171, 1989–1998.
Lichtenstein-Vidne, L., Henik, A., Safadi, Z., 2012. Task relevance modulates processing of distracting emotional stimuli. Cognit. Emot. 26, 42–52. https://doi.org/10.1080/02699931.2011.567055.
MacLeod, C., Mathews, A., Tata, P., 1986. Attentional bias in emotional disorders. J. Abnorm. Psychol. 95, 15. https://doi.org/10.1037/0021-843X.95.1.15.
Meier, S.M., Petersen, L., Mattheisen, M., Mors, O., Mortensen, P.B., Laursen, T.M., 2015. Secondary depression in severe anxiety disorders: a population-based cohort study in Denmark. Lancet Psychiatry 2, 515–523. https://doi.org/10.1016/S2215-0366(15)00092-9.
Meyer, T.J., Miller, M.L., Metzger, R.L., Borkovec, T.D., 1990. Development and validation of the Penn state worry questionnaire. Behav. Res. Ther. 28, 487–495. https://doi.org/10.1016/0005-7967(90)90135-6.
Miranda, R., Mennin, D.S., 2007. Depression, generalized anxiety disorder, and certainty in pessimistic predictions about the future. Cognit. Ther. Res. 31, 71–82. https://doi.org/10.1007/s10608-006-9063-4.
Nolen-Hoeksema, S., Morrow, J., 1991. A prospective study of depression and posttraumatic stress symptoms after a natural disaster: the 1989 Loma Prieta earthquake. J. Pers. Soc. Psychol. 61, 115–121. https://doi.org/10.1037/0022-3514.61.1.115.
Power, M., Dalgleish, T., 2015. Cognition and Emotion: from Order to Disorder, third ed. Psychology press, Hove and NY.
Priya, A., Garg, S., Tigga, N.P., 2020. Predicting anxiety, depression and stress in modern life using machine learning algorithms. Procedia Computer Science 167, 1258–1267.
Reiter, A.M., Atiya, N., Berwian, I., Huys, Q.J., 2020. Neuro-cognitive processes as mediators of psychological treatment effects. https://doi.org/10.31234/osf.io/wz5rd. (Accessed 5 October 2020).
Richter, T., Fishbain, B., Markus, A., Richter-Levin, G., Okon-Singer, H., 2020. Using machine learning-based analysis for behavioral differentiation between anxiety and depression. Sci. Rep. 10, 1–12. https://doi.org/10.1038/s41598- 020-72289-9.
Spielberger, C.D., Gorsuch, R.L., Lushene, R., Vagg, P.R., Jacobs, G.A., 1983. Manual for the State-Trait Anxiety Inventory. Consulting Psychologists Press, Palo Alto, CA.
Stehman, S.V., 1997. Selecting and interpreting measures of thematic classification accuracy. Remote Sens. Environ. 62, 77–89. https://doi.org/10.1016/S0034- 4257(97)00083-7.
Stuart, A.A., 1955. Test for homogeneity of the marginal distributions in a two-way classification. Biometrika 42, 412–416. https://doi.org/10.1093/biomet/42.3-4.412.
Tarsia, M., Power, M.J., Sanavio, E., 2003. Implicit and explicit memory biases in mixed anxiety–depression. J. Affect. Disord. 77, 213–225. https://doi.org/10.1016/S0165-0327(02)00119-2.
Van den Bergh, N., Marchetti, I., Koster, E.H., 2020. Bridges over troubled waters: mapping the interplay between anxiety, depression and stress through network analysis of the DASS-21. Cognit. Ther. Res. 1–15 https://doi.org/10.1007/s10608-020-10153-w.
Wu, Z., Fang, Y., 2014. Comorbidity of depressive and anxiety disorders: challenges in diagnosis and assessment. Shanghai Arch. Psychiatry 26, 227.