Article

Machine learning meta-analysis identifies individual characteristics moderating cognitive intervention efficacy for anxiety and depression symptoms

Check for updates

Thalia Richter [1,2,3,4,24] ✉, Reut Shani[1,2,3,24], Shachaf Tal[1], Nazanin Derakshan[5], Noga Cohen[6,7], Philip M. Enock[8], Richard J. McNally[8], Nilly Mor[9,10], Shimrit Daches[11], Alishia D. Williams[12], Jenny Yiend [13], Per Carlbring [14,15], Jennie M. Kuckertz[16,17,18], Wenhui Yang[19], Andrea Reinecke[20], Christopher G. Beevers[21], Brian E. Bunnell[22], Ernst H. W. Koster[23], Sigal Zilcha-Mano[1] & Hadas Okon-Singer[1,2,3]

Cognitive training is a promising intervention for psychological distress; however, its effectiveness has yielded inconsistent outcomes across studies. This research is a pre-registered individual-level meta-analysis to identify factors contributing to cognitive training efficacy for anxiety and depression symptoms. Machine learning methods, alongside traditional statistical approaches, were employed to analyze 22 datasets with 1544 participants who underwent working memory training, attention bias modification, interpretation bias modification, or inhibitory control training. Baseline depression and anxiety symptoms were found to be the most influential factor, with individuals with more severe symptoms showing the greatest improvement. The number of training sessions was also important, with more sessions yielding greater benefits. Cognitive trainings were associated with higher predicted improvement than control conditions, with attention and interpretation bias modification showing the most promise. Despite the limitations of heterogeneous datasets, this investigation highlights the value of large-scale comprehensive analyses in guiding the development of personalized training interventions.

Interventions targeting the cognitive mechanisms underlying anxiety and depressive symptoms have gained significant research interest in recent decades due to the high prevalence of these symptoms and the potential of online and remote treatment delivery.

Computerized cognitive interventions target the neurobiological mechanisms underlying cognitive and emotional domains by means of a demanding and engaging task that is repeatedly performed on a technological device[1,2]. Studies that have tested the efficacy of different cognitive training interventions in alleviating maladaptive symptoms of anxiety and depression have yielded contradictory outcomes (for a more thorough review see refs. 3,4).

Evidence suggests that specific characteristics of computerized cognitive training (such as the number of training sessions[5]) influence training efficacy. Moreover, some studies indicate that specific types of training methods are more effective in certain situations and/or for specific populations (e.g., greater improvements from attention bias modification for individuals dealing with generalized anxiety or depression, but lower improvements for those dealing with social anxiety and post-traumatic stress disorder[6]). Comparing the effectiveness of computerized training techniques, as well as the effect of individual differences on training efficacy, is highly valuable for several reasons: (a) they can provide guidelines to determine the optimal conditions and characteristics of effective training regimes[1]; (b) they can advance knowledge on central mechanisms underlying successful training[7]; and (c) they can promote individually tailored training programs[8]. Anxiety and depression are the most prevalent mental health disorders globally, imposing a substantial individual and economic burden that exceeds many other conditions[9,10]. Current treatments often yield only moderate success rates[11,12], partly due to their broad, non-personalized nature[13]. This underscores the urgent need for accessible, cost-effective interventions for anxiety and depression, that can be tailored to specific individual characteristics, to improve outcomes for those affected with maximum optimization.

A full list of affiliations appears at the end of the paper. ✉e-mail: trichter@campus.haifa.ac.il

Researchers have claimed that inconsistent outcomes across the cognitive training field may derive from methodological differences among cognitive training studies[14,15] as well as from individual differences in demographic, cognitive and clinical states[16]. More specifically, studies have found that the efficacy of training regimens can be affected by study-level characteristics, such as training duration[5] and training location[17], as well as by individual-level characteristics, such as age[18] and baseline cognitive ability[19]. In line with these findings it was proposed[19,20] that cognitive training interventions may be more beneficial for some individuals than for others and that group averages may conceal this insight.

These challenges and explanations are in line with a perspective article in *Nature*[21] whose authors suggest that treatment effects in behavioral science are invariably heterogeneous, rendering replication failures predictable. The authors challenge us to consider: "What if instead of treating variation in intervention effects as a nuisance or a limitation on the impressiveness of an intervention, we assumed that intervention effects should be expected to vary across contexts and populations?" (Page 982). They called for a "heterogeneity revolution" with the opportunity to investigate heterogeneity by considering effects of context and populations and accounting for moderating factors and subgroup differences, rather than relying on the assessment of average treatment effects.

In recent years, meta-analyses examining the efficacy of cognitive training for anxiety and depression have taken some steps toward identifying individual- or study-level moderators. However, their findings have been inconsistent, likely due to several factors: (a) a narrow focus on specific subsets of cognitive training, such as a single type of training, or isolated cognitive functions (e.g., working memory, inhibitory control) or cognitive biases (e.g., attention or interpretation bias modification); (b) reliance on aggregated study-level data rather than individual-level data; and (c) the use of classical statistical methods that struggle to address sample heterogeneity[4,6,22,23]. These studies highlight the dependency of the field on traditional approaches and limited datasets, which have left significant gaps in understanding the broader applicability and integration of cognitive training methods. These limitations call for approaches that can utilize individual-level data and advanced statistical methods capable of managing heterogeneity to investigate cognitive training[24].

Recent studies have highlighted the potential of machine learning (ML) methods to identify individual characteristics that influence treatment efficacy[25–28]. Due to their strong computational power, ML approaches may reveal specific patterns that might be missed by more traditional methods[29]. Rather than relying on statistical approaches that produce average effects, ML methods can consider the influence of non-linear and higher-order interaction effects, thus offering flexible modeling of moderating variables that can account for differences in training efficacy[30–34]. To our knowledge, our investigation is the first application of ML approaches in evaluating the contribution of individual- and study-level characteristics to the effectiveness of various cognitive training interventions for anxiety and depression symptoms.

Here, we attempt to move forward from the debate over the efficacy of cognitive training, in line with the suggestions of refs. 1,35 that a "one-size-fits-all" approach is unlikely to work. We do this by addressing individual- and study-level characteristics that may affect the efficacy of different cognitive training regimens. To that end, we designed a large-scale individual-level meta-analysis based on ML methods. Specifically, our study is a pre-registered multi-lab international endeavor designed to identify the prominent components underlying effective cognitive training interventions for anxiety and depression symptoms. We seek to identify the factors that influence the efficacy of several cognitive training types: (1) working

memory training, in which participants are asked to track and maintain goal-relevant information, despite the interference of distractions; (2) attention bias modification (ABM), in which participants implicitly learn to attend to neutral or positive stimuli while ignoring negative stimuli; (3) cognitive bias modification of interpretation (CBM-I), in which participants learn to interpret ambiguous information in a positive or neutral manner; and (4) inhibitory control training, in which participants are asked to inhibit irrelevant information[36,37].

We aimed to answer two questions: (Aim 1) What are the variables that affect training efficacy? (i.e., addressed by comparing training and control groups, irrespective of type of training); and (Aim 2) What are the variables that affect the efficacy of specific types of training? (addressed by comparing the four different training types).

Results

Please see Supplementary Information 1 for missing data information.

RF-based analysis, complete data set: Aim 1. Identifying who benefits from cognitive training

Table 1 presents the tuning parameters selected via the eightfold procedure, and the MSE values of the prediction in training and validation sets, for main and secondary outcomes.

As observed, the MSE values for the training and validation sets showed only minor differences across both outcomes, suggesting that the models are not overfitting to the training data. Additionally, both outcome predictions demonstrate relatively high accuracy, with MSE values falling within less than one standard deviation of the standardized outcome scale.

Figures 1 and 2 show the importance of the features and SHAP values of the prediction of the main outcome.

As observed from Figs. 1 and 2, the most contributing moderators were the baseline score of the main outcome measure and number of training sessions. The SHAP values indicate that participants with higher baseline anxiety and depression symptoms showed more improvement and that an increased number of training sessions led to more improvement.

In addition, the optimal condition (i.e., training or control) was predicted for each participant based on the trained and tested RF model. Table 2 shows the number of participants expected to be assigned to each group.

As shown in Table 2, 838 participants (62%) were predicted to achieve greater gain when assigned to one of the training groups.

Figures 3 and 4 show the importance of the features and SHAP values of the prediction of the secondary outcome.

As shown by Figs. 3 and 4, the moderators with the greatest contribution were the aggregated baseline score of the secondary outcome measure and the training session duration. The SHAP values indicate that participants with higher baseline anxiety and depression symptoms improved more and that a medium session duration was the most effective for improvement.

In addition, the optimal condition (i.e., training or control) was predicted for each participant based on the trained and tested RF model. Table 3 shows the number of participants expected to be assigned to each group.

As shown in Table 3, 799 participants (71%) were predicted to achieve greater gain when assigned to one of the training groups.

RF-based analysis, complete data set: Aim 2. Identifying who benefits from a specific type of training

Table 4 presents the tuning parameters selected via the 8-fold procedure, and the MSE values of the prediction in training and validation sets, for the main and secondary outcomes.

Table 1 | Selected tuning parameters and MSE values for the main and secondary outcomes prediction of Aim 1

Outcome	N trees	Minimum samples split	Minimum samples leaf	Max features	Max depth	Training MSE	Validation MSE
Main	25	9	2	4	6	1.08	0.92
Secondary	10	12	8	5	21	0.43	0.46

Fig. 1 | Feature importance of Aim 1: main outcome. The relative contribution of each moderator to the prediction. The zero line represents zero importance. *Note:* pre_main - baseline score in the main outcome, nsessions—number of training sessions, country—international or USA; population_1/2/3—anxious, depressed, or healthy, accordingly; diagnosis—diagnosis method: clinical assessment or self-report questionnaire; at_home—training conducted online or at the lab; is training—training or control conditions, has_emotionl_stimuli – emotional stimuli were present or absent in the training.
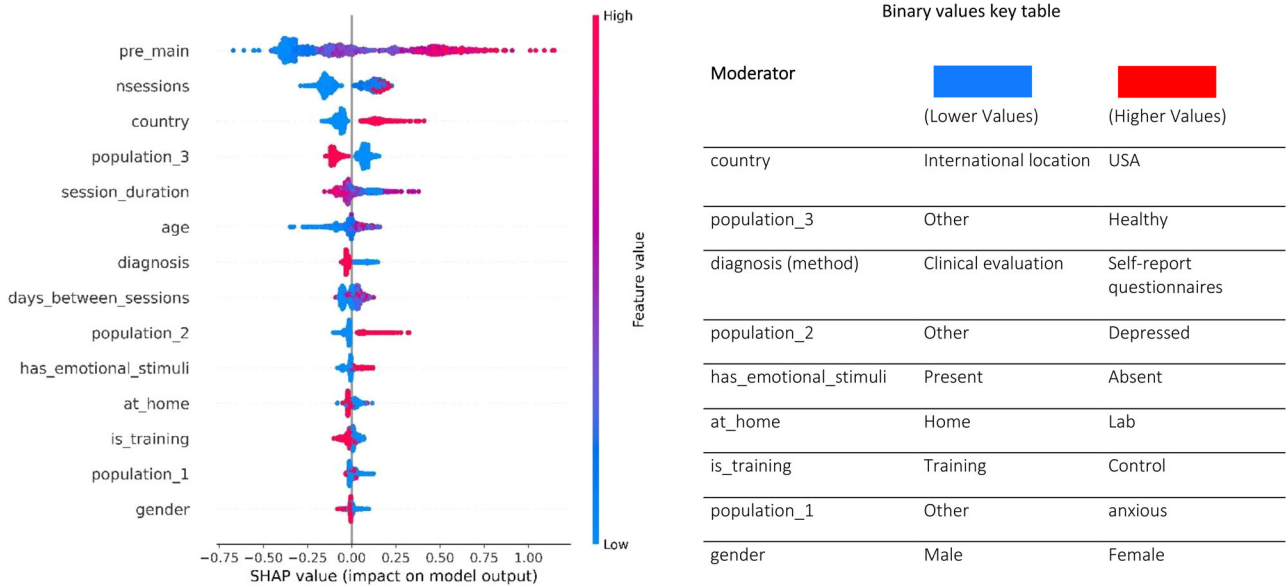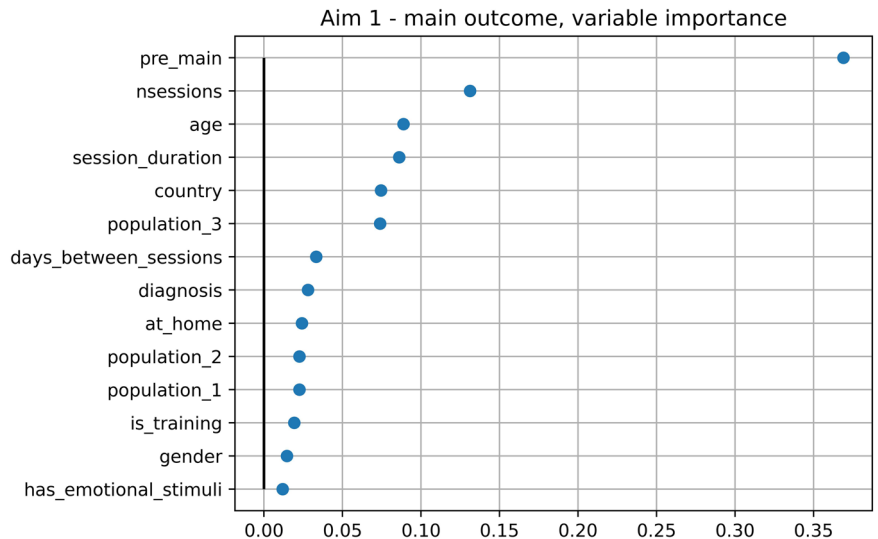

Fig. 2 | SHAP values for Aim 1: main outcome. SHAP values for each moderator and the corresponding binary value key table. Features with positive SHAP values contribute positively to the outcome, while those with negative values have a negative effect. Blue indicates lower values for each moderator, whereas red indicates higher values. When the values are binary, the original coding determines classification to "lower" and "higher". For instance, the "diagnosis" (method) feature was coded as "1" for clinical evaluations and "2" for self-report questionnaires; therefore, clinical evaluations are shown in blue, and self-report questionnaires are in red. Thus, the graph indicates that participants with clinical diagnoses are expected to gain more improvement than those with sub-clinical symptom levels. However, this moderator's overall importance for prediction is minimal, as indicated in Fig. 1. For continuous values, color coding is also continuous without clear cut-offs. See Fig. 1 for the abbreviations of all moderators.

Table 2 | Optimal group prediction for Aim 1: main outcome

		Original group		
		Training	**Control**	**Total**
Optimal group	Training	452	386	838
	Control	242	253	495
	Total	694	639	1333

As observed, the MSE values for the training and validation sets showed only minor differences across both outcomes, suggesting that the models are not overfitting to the training data. Additionally, both outcome predictions demonstrate relatively high accuracy, with MSE values falling within less than one standard deviation of the standardized outcome scale.
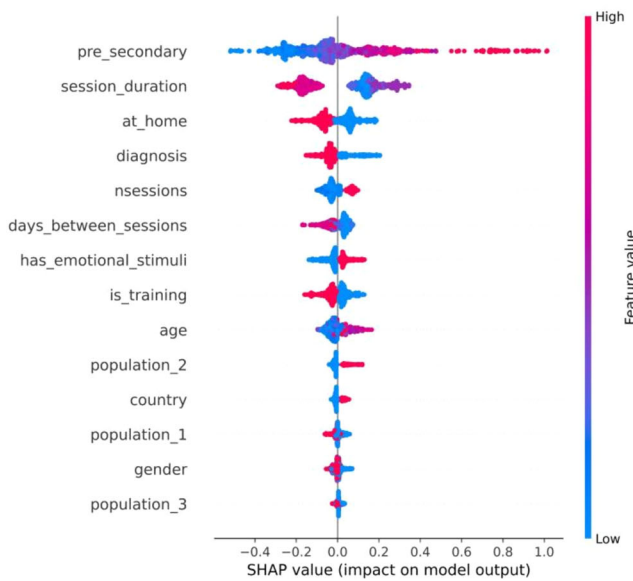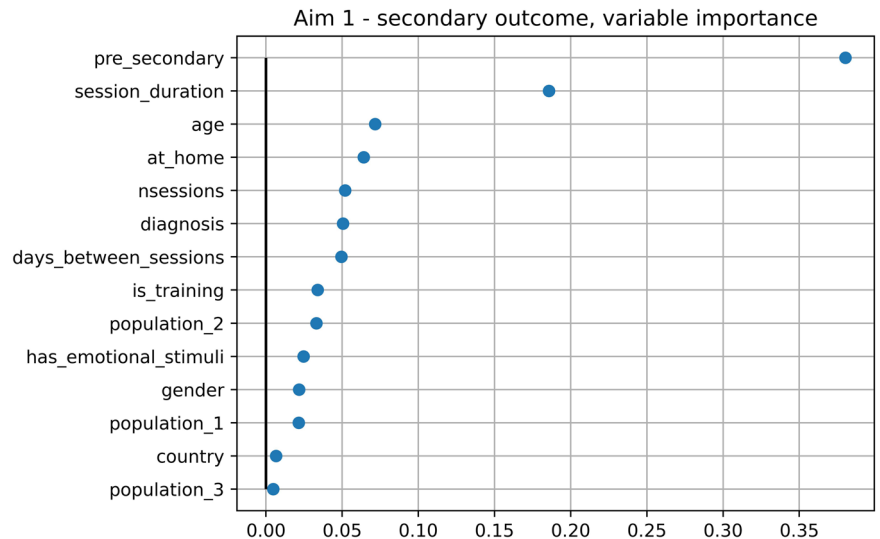
Figures 5 and 6 show the importance of the features and SHAP values of the prediction of the main outcome.

As shown in Fig. 5 and 6, the key contributing moderators were the baseline score of the main outcome measure, followed by the number of sessions. The SHAP values suggest that participants with higher baseline anxiety and depression symptoms showed more improvement and that an increased number of sessions led to greater gain.

In addition, the optimal training condition (i.e., working memory training, ABM, CBM-I, or inhibitory control training) was predicted for each participant based on the trained and tested RF model. Table 5 shows the number of participants expected to be assigned to each group.

As shown in Table 5, 557 participants (80%) were predicted to achieve greater gain when assigned to ABM, 134 participants (19%) to CBM-I, 3 participants to inhibitory control training, and no participants were predicted to gain from working memory training.

Fig. 3 | Feature importance of Aim 1: secondary outcome. The relative contribution of each moderator to the prediction. The zero line represents zero importance. *Note:* pre_secondary= baseline score in the secondary outcome. See Fig. 1 for the abbreviations of all moderators.

Fig. 4 | SHAP values for Aim 1: secondary outcome. SHAP values for each moderator and the corresponding binary value key table. Features with positive SHAP values contribute positively to the outcome, while those with negative values have a negative effect. Blue indicates lower values for each moderator, whereas red indicates higher values. When the values are binary, the original coding determines classification to 'lower' and 'higher'. *Note:* pre_secondary= baseline score in the secondary outcome. See Fig. 1 for the abbreviations of all moderators.

Table 3 | Optimal group prediction for Aim 1: Secondary outcome

		Original group		
		Training	**Control**	**Total**
Optimal group	Training	406	393	799
	Control	182	135	317
	Total	588	528	1116

Figures 7 and 8 show the importance of the features and SHAP values of the prediction of the secondary outcome.

As observed from Figs. 7 and 8, the most contributing moderators were the aggregated baseline score of the secondary outcome measure and the duration of the training session. The SHAP values indicate that participants with higher baseline anxiety and depression symptoms improved more following training and that medium and short session durations were more effective for improvement than long ones.

In addition, the optimal training condition (i.e., working memory training, ABM, CBM-I, or inhibitory control training) was predicted for each participant based on the trained and tested RF model. Table 6 shows the number of participants expected to be assigned to each group.

As shown in Table 6, 108 participants (18%) were predicted to achieve greater gain when assigned to ABM, 480 participants (82%) to CBM-I, and no participants were predicted to gain from working memory or inhibitory control training.

LME analysis, clinical data set: Aim 1. Identifying who benefits from cognitive training

The linear mixed effects (LME) model was constructed twice for the clinical data set, while after the first estimation, all non-significant interactions were removed.

Table 4 | Selected tuning parameters and MSE values for the main and secondary outcomes for prediction of Aim 2

Outcome	N trees	Minimum samples split	Minimum samples leaf	Max features	Max depth	Training MSE	Validation MSE
Main	50	2	8	5	19	1.20	0.83
Secondary	100	4	7	9	8	0.43	0.41

Fig. 5 | Feature importance of Aim 2: main outcome. The relative contribution of each moderator to the prediction. The zero line represents zero importance. *Note:* training_type_1/2/3/4= working memory training, ABM, CBM-I, inhibitory control training, accordingly. See Fig. 1 for the abbreviations of all moderators.
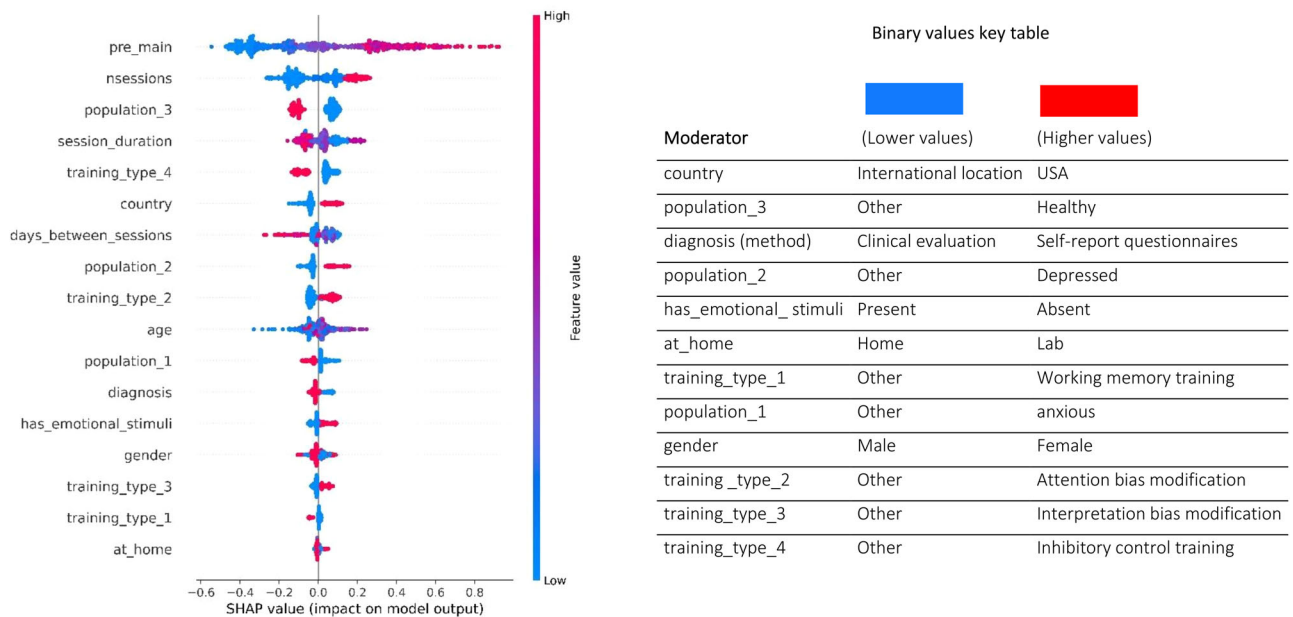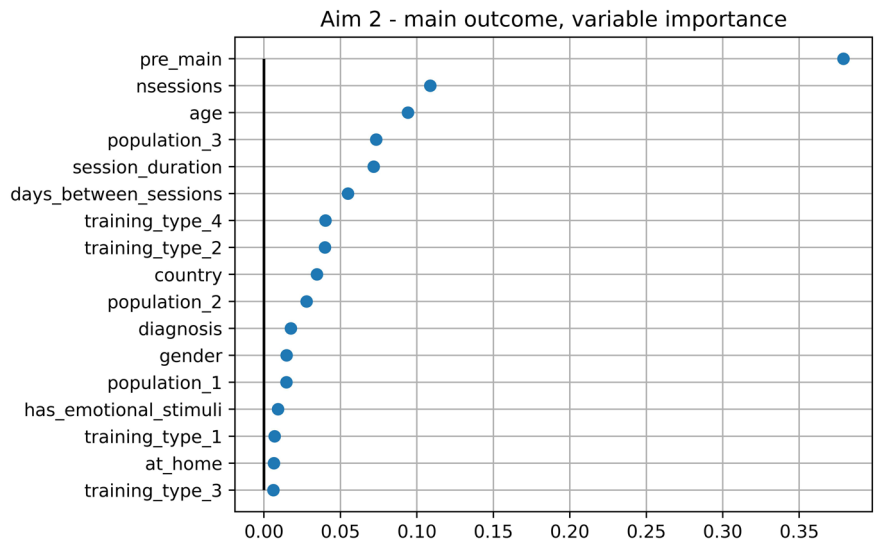

Aim 2 - main outcome, variable importance

Fig. 6 | SHAP values for Aim 2: main outcome. SHAP values for each moderator and the corresponding binary value key table. Features with positive SHAP values contribute positively to the outcome, while those with negative values have a negative effect. Blue indicates lower values for each moderator, whereas red indicates higher values. When the values are binary, the original coding determines classification to "lower" and "higher". See Fig. 1 for the abbreviations of all moderators.

Table 5 | Optimal group prediction for Aim 2: Main outcome

		Original group				
		Working memory training	ABM	CBM-I	Inhibitory control training	Total
Optimal group	Working memory training	0	0	0	0	0
	ABM	72	168	118	199	557
	CBM-I	0	72	50	12	134
	Inhibitory control training	0	2	0	1	3
	Total	72	242	168	212	694

Fig. 7 | Feature importance of Aim 2: secondary outcome. The Relative contribution of each moderator to the prediction. The zero line represents zero importance. *Note:* pre_secondary= baseline score in the secondary outcome. training_type_1/2/3/4= working memory training, ABM, CBM-I, inhibitory control training, accordingly See Fig. 1 for the abbreviations of all moderators.
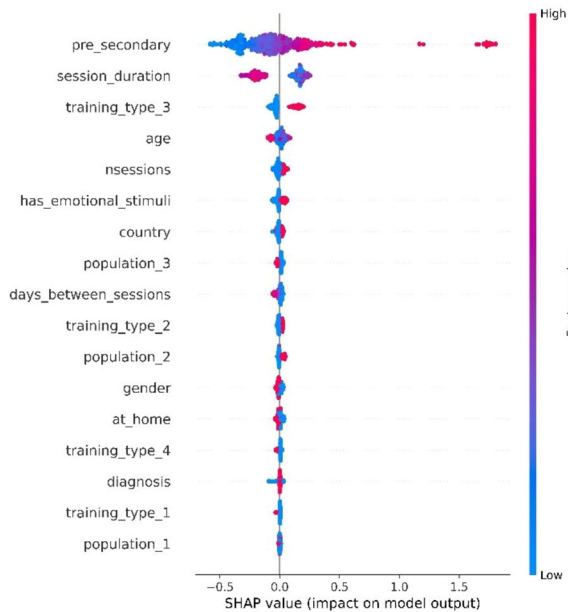
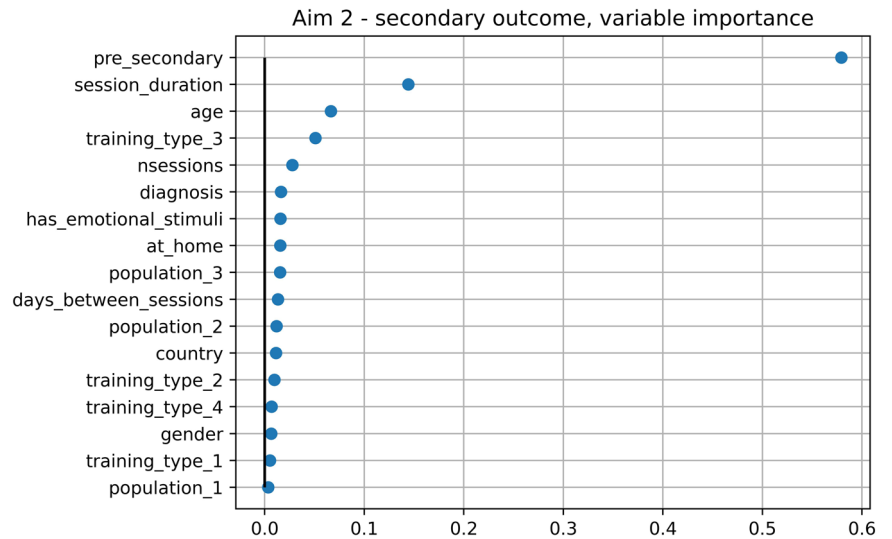

Fig. 8 | SHAP values for Aim 2: secondary outcome. SHAP values for each moderator and the corresponding binary value key table. Features with positive SHAP values contribute positively to the outcome, while those with negative values have a negative effect. Blue indicates lower values for each moderator, whereas red indicates higher values. When the values are binary, the original coding determines classification to "lower" and "higher". See Fig. 1 for the abbreviations of all moderators.

Table 6 | Optimal group prediction for Aim 2: Secondary outcome

		Original group				
		Working memory training	**ABM**	**CBM-I**	**Inhibitory control training**	**Total**
Optimal group	Working memory training	0	0	0	0	0
	ABM	5	102	1	0	108
	CBM-I	60	140	136	144	480
	Inhibitory control training	0	0	0	0	0
	Total	65	242	137	144	588

Table 7 presents the MSE values of the prediction for the main and secondary outcomes.

The MSE values of both outcomes fall within less than 1.5 standard deviations of the standardized outcome scale.

Table 8 presents the Beta coefficients of the moderators and interactions found significant in the first estimation and inserted into the second estimation.

As shown in Table 8, a main effect for the baseline main outcome was found ($\beta = 0.45, t(158) = 5.97, p<.001$) such that higher baseline anxiety and depression symptoms were associated with more gain. A main effect for gender was also found ($\beta = -0.53, t(158) = -2.59, p<0.05$), such that women had lower gains than men. In addition, an interaction effect was found between the control condition and the depression and anxiety and other conditions. Multiple comparisons with Bonferroni correction

Table 7 | MSE values for the main and secondary outcomes prediction of Aim 1

	MSE
Main outcome	1.32
Secondary outcome	1.49

Table 8 | Beta coefficients of Aim 1: main outcome

Predictors	Estimates	Gain main CI	p
(Intercept)	0.28	-0.92–1.48	0.643
pre main	0.45	0.30–0.59	**<0.001**
age	-0.01	-0.02–0.01	0.515
gender [1]	-0.53	-0.93– -0.13	**0.011**
Current medications [1]	-0.22	-0.62–0.18	0.270
nsessions	0.07	-0.01–0.15	0.081
is training [2]	-0.05	-0.67–0.58	0.886
Type of diagnosis [1]	-0.03	-0.59–0.54	0.928
Type of diagnosis [2]	0.2	-0.97–1.36	0.737
Type of diagnosis [3]	-1.43	-2.59– -0.27	**0.016**
is training [2] × Type of diagnosis [1]	-0.39	-1.18–0.39	0.324
is training [2] × Type of diagnosis [2]	-0.21	-2.94–2.52	0.878
is training [2] × Type of diagnosis [3]	4.51	1.83–7.19	**0.001**
Random Effects			
σ^2	1.43		
τ_{00} study	0.00		
N study	3		
Observations	171		
Marginal R^2/Conditional R^2 NA	0.356		

Beta coefficients, confidence interval, and p-value of all moderators and interactions in the second estimation of the LME model for the clinical data set. Note: pre_main - baseline score in the main outcome, gender [1] - women, Current medications [1] - yes, is training [2] - control, Type of diagnosis [0/1/2/3] – depression; depression and anxiety; depression and other conditions; depression and anxiety and other conditions, respectively, nsession - number of sessions.

Table 9 | Beta coefficients of Aim 1: secondary outcome

Predictors	Estimates	Gain secondary CI	p
(Intercept)	1.14	-10.01–12.29	0.840
age	0	-0.02–0.02	0.892
gender [1]	-0.56	-1.03– -0.09	**0.020**
Current medications [1]	-0.43	-0.90–0.05	0.078
nsessions	-0.17	-1.64–1.30	0.816
Type of diagnosis [1]	0.24	-0.29–0.77	0.376
Type of diagnosis [2]	0.09	-1.11–1.29	0.882
Type of diagnosis [3]	0.31	-0.89–1.52	0.606
is training [2]	0.31	-0.40–1.02	0.391
pre secondary	0.61	0.42–0.80	**<0.001**
is training [2] × pre secondary	-0.25	-0.48– -0.03	**0.029**
Random Effects			
σ^2	1.62		
τ_{00} study	0.23		
ICC	0.12		
N study	3		
Observations	144		
Marginal R^2/Conditional R^2	0.394/0.470		

Beta coefficients, confidence interval, and p-value of all moderators and interactions in the second estimation of the LME model for the clinical data set. Note: pre_secondary—baseline score of the secondary outcome. See Table 8 for all abbreviations and value keys.

Table 10 | MSE values for the main and secondary outcomes prediction of Aim 2

	MSE
Main outcome	1.31
Secondary outcome	1.29

revealed that in the control condition, participants with depression, anxiety, and other psychiatric conditions showed significantly greater gains than participants with only depression and anxiety ($\beta = -3.50$, $t(158) = -2.86$, $p < 0.05$).

Table 9 presents the Beta coefficients of the moderators and interactions found significant in the first estimation and inserted into the second estimation.

As shown in Table 9, a main effect for the baseline secondary aggregated outcome was found ($\beta = 0.61$, $t(132.87) = 6.37$, $p < .001$), such that higher anxiety and depression symptoms were associated with more gain. A main effect for gender was also found ($\beta = -0.56$, $t(132.04) = -2.35$, $p < 0.05$), such that women had lower gains than men. Finally, an interaction effect between the control condition and the baseline secondary aggregated outcome was found ($\beta = -0.25$, $t(132.45) = -2.21$, $p < 0.05$), such that the effect of the baseline score on gain is weaker in the control condition.

LME analysis, clinical data set: Aim 2. Identifying who benefits from a specific type of training

Due to almost complete multicollinearity between the training type and the moderators "number of sessions" and "type of diagnosis", these two

moderators were not included in the model. Table 10 presents the MSE values of the prediction for main and secondary outcomes.

The MSE values of both outcomes fall within less than 1.5 standard deviations of the standardized outcome scale.

Table 11 presents the Beta coefficients of the moderators and interactions found significant in the first estimation and inserted into the second estimation.

As shown in Table 11, no significant interactions were found in the first estimation of the model. Therefore, no interactions were inserted into the second model. After the second estimation, a main effect for the baseline main outcome was found ($\beta = 0.42$, $t(80) = 3.319$, $p < 0.05$), such that a higher baseline score was associated with more gain.

Table 12 presents the Beta coefficients of the moderators and interactions found significant in the first estimation and inserted into the second estimation.

As shown in Table 12, a main effect for the baseline secondary aggregated outcome was found ($\beta = 0.67$, $t(77.28) = 6.41$, $p < .001$), such that a higher baseline score was associated with more gain. In addition, the interaction effect between the CBM-I training and the medication status was significant. Multiple comparisons with Bonferroni correction revealed that participants not taking medications showed significantly greater improvement in CBM-I training compared to those who were on medications ($\beta = 0.98$, $t(78.02) = 3.15$, $p < 0.05$).

Table 11 | Beta coefficients of Aim 2: Main outcome

Predictors	Estimates	Gain main CI	p
(Intercept)	0.84	-0.52–2.20	0.224
training type [3]	0.09	-0.69–0.88	0.811
pre main	0.42	0.17–0.68	**0.001**
age	-0.01	-0.04–0.01	0.357
gender [1]	-0.39	-0.96–0.17	0.170
Current medications [1]	-0.1	-0.66–0.47	0.734
Random Effects			
σ²;	1.41		
τ00 study	0.00		
N study	3		
Observations	86		
Marginal R²/ Conditional R²	0.187/NA		

Beta coefficients, confidence interval, and p-value of all moderators and interactions in the second estimation of the LME model for the clinical data set. Note: training type [2/3] - attention bias modification or interpretation bias modification, respectively. See Table 8 for all abbreviations and value keys.

Table 12 | Beta coefficients of Aim 2: Secondary outcome

Predictors	Estimates	Gain secondary CI	p
(Intercept)	-0.53	-1.77–0.70	0.393
Pre secondary	0.67	0.46–0.88	**<0.001**
age	0.01	-0.02–0.03	0.577
gender [1]	-0.46	-1.01–0.10	0.109
training type [3]	0.24	-1.21–1.68	0.746
Current medications [1]	0.52	-0.79–1.83	0.431
Training type [3] × Current medications [1]	-1.51	-2.94– -0.07	**0.040**
Random Effects			
σ²	1.42		
τ00 study	0.17		
ICC	0.11		
N study	3		
Observations	86		
Marginal R²/ Conditional R²	0.514/0.567		

Beta coefficients, confidence interval, and p-value of all moderators and interactions in the second estimation of the LME model for the clinical data set. Note: pre_secondary - baseline score of the secondary outcome. training type [2/3] - attention bias modification or interpretation bias modification, respectively. See Table 8 for all abbreviations and value keys.

Discussion

Cognitive training interventions are being developed and tested worldwide in both academic and commercial settings, as they demonstrate promising, though inconsistent, potential for improving cognitive and emotional capacities and/or efficiency[38,39]. The current study is the first to evaluate the contribution of study-level (i.e., training characteristics) and individual-level (i.e., demographic and symptomatic characteristics) moderators to the efficacy of cognitive training for anxiety and depression symptoms and of certain training types in particular. To this end, we implemented ML methods to analyze individual-level data collected from 1544 participants across the globe, all of whom completed one of the following cognitive training interventions or its corresponding control activity: working memory training, attention bias modification, interpretation bias modification, and inhibitory control training.

To investigate factors that moderate the efficacy of cognitive training for anxiety and depression symptoms, we grouped all cognitive training types to identify those who benefit from cognitive training in general and compared their efficacy to all control groups. The RF model exhibited a relatively high accuracy for predicting gain, as reflected in the MSE values, in two analyses in which the main outcome measure and a secondary aggregated outcome measure were applied. The feature importance procedure showed which individual and study-level characteristics contribute to the prediction of training efficacy. *Baseline outcome score* and *training dose* (number of sessions in the main outcome analysis and session duration in the secondary outcome analysis) emerged as the most important moderators for predicting training efficacy. The SHAP values indicated that participants with elevated baseline scores improved more than those with lower baseline scores. An increased number of sessions and medium duration of sessions were also associated with more improvement following training. In our dataset, the number of training sessions ranged from 1 to 15, with only one study having 84 sessions, and session durations varied between 2 and 30 min. Since the SHAP values do not specify cut-offs for color coding, the specific number and duration of sessions that were associated with the highest improvement cannot be directly inferred. Additionally, the training condition resulted in greater gains compared to the control condition. Accordingly, the training condition was predicted to be the optimal group for most participants.

The second aim was to identify who benefits from each specific type of cognitive training. Once again, the RF model yielded relatively accurate predictions for the main and secondary outcomes, as reflected in the MSE values. When examining feature importance, the most important moderators were again *baseline outcome score* and *training dose* (number of sessions in the main outcome analysis and session duration in the secondary outcome analysis). Similarly to the results of Aim 1, the SHAP values indicated that participants with increased baseline scores improved more and that an increased number of sessions, and medium or short duration of sessions, led to more improvement following training. For the main outcome, ABM and CBM-I were associated with more gains, while working memory and inhibitory control training were not. The optimal group predicted for most of the participants was the ABM group. For the secondary outcome, the CBM-I group was associated with more improvement than the other groups, and in accordance, it was predicted as the optimal group for most of the participants.

When focusing exclusively on participants with a clinical diagnosis, who represent a highly relevant population for cognitive training interventions, only an LME model could be employed due to the limited number of participants with available clinical data. The model demonstrated sound predictive accuracy across primary and secondary outcomes. When analyzing individual- and study-level moderators that might influence training efficacy among patients, we found that patients with more severe symptoms and men, more than women, exhibited greater improvements. The effect of the baseline score was attenuated in the secondary outcome, such that the severity of baseline symptoms was less predictive of training efficacy in the control condition. Additionally, patients with multiple diagnoses showed significantly greater improvement in the control condition, suggesting that training and control conditions may not be distinctly different for this group. Finally, when considering moderators in the context of comparing different types of training, the baseline score effect was consistent for both main and secondary outcomes. Furthermore, medication status influenced gains in the CBM-I training, with non-medicated patients showing greater improvement.

Overall, as we detail below, our results point to specific factors that modulate the efficacy of cognitive training. The results suggest that individual and training characteristics play a central role. At the individual level, the individual's clinical status (represented by a clinical diagnosis or baseline score on the selected self-report mental health questionnaire) is a fundamental factor in determining the success of cognitive training. This is

somewhat aligned with the results published[6] that reported moderating significant effects for "clinical status" (clinical or non-clinical sample) in CBM-I and "symptom type" (type of mental disorder) in ABM.

At the training level, dose-related ingredients of the training (e.g., number of training sessions and session duration) were pivotal to training efficacy. This result is consistent with previous meta-analytical reviews that focused on older adults[40] and on brain-injured patients[41]. Elsewhere, however, for example in schizophrenia research, evidence for the moderating role of baseline symptom severity and training dose in cognitive training efficacy has been mixed. While several meta-analyses[42–44] found no moderating role for baseline symptom severity, others[45] found that the higher the symptoms the better the outcome of training. Similar mixed results were reported for symptom duration[45,46]. These studies relied on classical statistical methods and primarily aggregated study-level data, which may have contributed to the conflicting findings. In our study, the ML-based framework addresses these limitations by leveraging individual-level data and integrating multiple cognitive training modalities, providing a promising path for resolving such inconsistencies in schizophrenia research and other conditions beyond anxiety and depression.

Although men improved more than women in the LME model of the complete and clinical data sets, this effect was not attenuated by condition (training vs. control), and the gender moderator showed only minimal contribution in the RF-based model. Therefore, it seems that both age and gender had a minimal influence in predicting improvement values. This finding is in line with a recent systematic review that examined the effects of cognitive control training on anxiety and depression in children and adolescents. This review revealed that the impact of age on changes in anxiety and depression symptoms is nonsignificant[47]. Another meta-analytic review that focused on depressed individuals found a moderating effect of age but not gender[38]. A review of six meta-analyses on the efficacy of CBM-I revealed that most studies found no significant effects of gender or age[48]. In the current data, age was distributed evenly in both groups, and therefore, it is less plausible that the small contribution of this moderator to the prediction model is due to unbalanced distribution (e.g., mean age was 29.3, SD = 11.2, for the training group, and 28.2, SD = 10 for the control group). Although the participants in the meta-analysis ranged in age from 18 to 74, the majority were younger (18–40). Therefore, the minimal impact of age should be considered with caution beyond this range. Future studies comparing the efficacy of cognitive training across different age groups are needed to determine whether significant age differences affect training outcomes.

Our second analysis, which sought to identify subgroups of individuals who benefit from one type of cognitive training more than other types, indicated that ABM and CBM-I may lead to greater improvements for certain participants compared to working memory or inhibitory control training. This extends previous studies that have demonstrated the effectiveness of these methods across various psychiatric disorders and diverse populations[6]. These cognitive trainings have also shown far-transfer effects, meaning they not only reduce the biases themselves but also alleviate associated psychopathological symptoms, such as depression and anxiety[6].

Previous evidence supports that CBM-I could be particularly beneficial for individuals with more severe symptoms. In a recent review of CBM-I, sample type, and associated baseline symptom level, influenced the effectiveness of CBM-I training[49]. Related to this, a recent study investigating the association of different cognitive biases to anxiety and depression[50] found that interpretation biases were superior in predicting both anxiety and depression symptoms severity, highlighting the potential advantage of CBM-I in these contexts.

The current study represents an innovative attempt to investigate several cognitive training methods based on an individual-level synthesized dataset in the field of anxiety and depression. We must be mindful that the extensive heterogeneity of the included studies in terms of training characteristics, outcome measures, and populations is a potential limitation. It is possible that some moderator effects did not emerge due to excessive heterogeneity, and that individual differences were obscured by outcome

standardization. However, the study highlights that to pursue large-scale meta-analytic studies, it is important to develop methods to unify outcome measures that will facilitate the comparison among studies. Here, we standardized various questionnaires based on their respective normative data. An alternative is to establish gold-standard questionnaires for outcome measures. Efforts to designate questionnaires as common outcome measures for mental health research have been led by health organizations such as the Schizophrenia International Research Society[51], Wellcome[52], and the International Alliance of Mental Health Research Funders (IAMHRF). In the context of depression and anxiety, the IAMHRF recommends the use of the patient health questionnaire (PHQ-9), the generalized anxiety disorder assessment (GAD-7), and the world health organization disability assessment schedule- 2.0 (WHODAS) as standard outcome measures for comparison and integration between disorders and efficacy studies[53].

Another solution, recently suggested by Perlman and colleagues[54], is to categorize questions into a detailed taxonomy, standardizing responses, and transforming variables to ensure consistency across studies. For instance, questions on functional impairment from different questionnaires can be grouped under the "quality of life" category with a subcategory for "functional impairment". This approach allows similar questions to be combined into a unified feature, which can then be used in predictive models or as standard outcomes. In their study, the researchers synthesized data from 17 studies to create a model predicting differential treatment benefits for six potential medication options, successfully forecasting treatment efficacy.

Our rich and varied dataset may have posed other limitations. First, since our groups were composed of individual-level data collected at different sites and using different methodological practices, the distribution of moderating factors was not always balanced across studies. This is reflected by differences between the groups (e.g., some training groups contained more participants with anxiety or depression than other groups), potentially distorting some of the findings. To minimize group differences, we did not include potential moderators that were insufficiently represented in all training groups (e.g., the adaptive nature of the training). Other meta-analyses have faced similar challenges when analyzing heterogeneous data with different types of cognitive training interventions. For example, a recent study[1] took a different approach by merging five cognitive training categories into two groups (single-domain training and multi-domain training) due to insufficient representations for each type of training. This example highlights the tradeoffs necessary when synthesizing an array of datasets.

Heterogeneous datasets present significant challenges, including difficulty harmonizing diverse measures and the increased statistical complexity of identifying moderators within heterogeneous groups. However, this variability mirrors the real-world diversity inherent in cognitive training applications, making the findings potentially more robust and widely generalizable across populations. Addressing such variability is essential for conducting large-scale, individual-level meta-analyses that can deliver actionable and broadly applicable insights.

ML provides a powerful advantage by allowing for more robust validation and testing of models on diverse datasets. One of its inherent strengths lies in its ability to handle complex associations within the data that may be camouflaged and remain undetected if conventional statistical methods are used, thus offering a more nuanced and deeper analysis of outcomes. By setting aside a portion of the data for testing, ML ensures that models are validated on unseen data, enhancing their generalizability to broader populations and outcomes. This rigorous process addresses the limitations often seen with mixed findings in traditional methods, ensuring more reliable results across varying settings. Consequently, ML models may yield stronger, more consistent predictions, particularly in studies involving heterogeneous settings and complex variables.

Additionally, while SHAP values provided some level of interpretability to the RF results by offering directional insights into the predictions, they still represent a simplified approach to presenting the complex patterns and associations detected among the moderators within the RF model. Future research should take advantage of the recent exciting advances in the

field of artificial intelligence. These advances promote the development of explainable artificial intelligence models that are expected to provide more accurate predictions in addition to revealing possible causal mechanisms. For an elaborated review and discussion, see refs. 55–57.

The present study provides the first comprehensive ML-based individual-level meta-analysis of various cognitive training targeting anxiety and depression symptoms. Our work addresses the need to detect individual characteristics that moderate training efficacy by advancing two critical aims: (1) identifying individual characteristics that predict who may benefit most from cognitive training, and (2) comparing the efficacy of different types of cognitive training interventions.

Our study uniquely integrates ML methods to analyze a harmonized, individual-level dataset compiled from diverse studies, addressing key limitations associated with data heterogeneity. This study advances the existing literature in several important ways:

(a) It is the first meta-analysis to comprehensively compare cognitive function training (e.g., working memory, inhibitory control) and cognitive bias modification (CBM) approaches (e.g., attention bias modification, interpretation bias modification) across four distinct types of training for anxiety and depression symptoms.

(b) By harmonizing data from multiple studies with diverse outcome measures, we introduce a possible pipeline for managing heterogeneous datasets. This methodology offers an applicable framework for future research facing similar challenges.

(c) While the use of ML to identify individual characteristics that inform treatment efficacy was previously studied, our application to cognitive training interventions for anxiety and depression offers a meaningful and insightful contribution on several levels. Our approach addresses specific challenges in this domain, such as handling directional effects, leveraging a large dataset, and aggregating individual-level data. By identifying both individual- and study-level characteristics that moderate treatment outcomes, our study provides critical insights that pave the way for more tailored and effective interventions. Building on our findings, personalized treatments for anxiety and depression should prioritize participants' baseline emotional and mental health status to improve efficacy and incorporate a greater number of training sessions to maximize benefits.

Methods
Due to the complexity of our analysis, we began by publishing a detailed, preregistered protocol[8]. This protocol describes the methods of data collection, data processing, main and secondary outcomes, potential moderators, and planned analysis, as well as the rationale behind these steps. In the following section, we provide a summary of our methods.

Data collection
A comprehensive PubMed literature search was conducted, generating 574 articles published from 2013 to 2018, which were reviewed by two experienced team members. The studies were examined for adherence to our inclusion-exclusion criteria: healthy, sub-clinical, or anxious and depressed adults between the ages of 18 and 75; targeting one specific cognitive function (e.g., working memory, attention, interpretation, inhibition); and including standardized questionnaires as outcome measures. The corresponding authors of the 39 eligible papers were contacted and asked to contribute their individual-level datasets, stripped of identifying characteristics. We received 22 datasets comprising 1544 participants[58–78]. All participants had completed one of the following procedures: working memory training (5 datasets), attention bias modification training (ABM) (6 datasets), cognitive interpretation bias modification (CBM-I) (5 datasets), inhibitory control training (6 datasets), or a control condition corresponding to one of these training types.

Standardization of outcome measures
Since each dataset in the study included different questionnaires for outcome measures, standardizing the scores was necessary to compare

outcomes across studies. To achieve this, we conducted a literature search to identify healthy population norms (i.e., means and standard deviations) for each questionnaire. When possible, the country in which the population norms were collected was matched to the country where each study was conducted. Standardization was performed by subtracting the population mean from each participant's score and dividing it by the standard deviation. This method, which accounted for the varying scales of the different questionnaires, allowed us to calculate a standardized difference score for each participant. As a result, all outcome measures were unified into a single comparable scale, facilitating consistent comparisons. For further discussion on the importance of pre-registered and reliable outcome measures, see ref. 79.

To evaluate the efficacy of each training, we calculated two outcome measures for each study:

Main outcome measure
To establish main outcomes prior to analysis, we provided authors with the following guidelines for selecting a primary outcome measure for each included study: (1) The variable should be a specific score from a standardized questionnaire related to mental health, emotions, or well-being, and must have corresponding healthy population norms for comparability; (2) The chosen variable should align with the primary focus outlined by the investigators in their original published work. All outcome measures represent the difference between pre- and post-cognitive training scores on a standardized questionnaire.

Secondary outcome measure
Secondary outcome scores were calculated by aggregating and averaging scores on all outcome measures not selected as a main outcome in the current study (i.e., standardized questionnaires which were collected before and after the training but were not selected as the main outcome score in the current study). For each study, standardized outcome scores for all reported secondary measures were aggregated and averaged to produce a single outcome score representing the difference between pre- and post-training (see protocol for further details).

Potential moderators
We examined both individual-level and study-level moderators. The individual-level moderators included age, gender (male or female), and baseline standardized main and secondary outcome scores (respective to the outcome measure used in each analysis). The study-level moderators were population type (depressed, anxious, or healthy; respective to the inclusion criteria of the original study), clinical assessment method (clinical assessment or self-report questionnaire), number of cognitive training sessions, average number of days between cognitive training sessions, average cognitive training session duration (in minutes), geographical region (US or international), cognitive training location (lab or online), the inclusion of a visual emotional stimulus (present or absent). Please see the protocol for further details on each moderator. In a separate analysis not included in the original protocol, the following clinical individual-level moderators were examined as well: medication status (whether currently undergoing medications or not), and type of diagnosis (depression; depression and anxiety; depression and other conditions; depression and anxiety and other conditions).

Data analysis. Figure 9 shows a summary of the analysis workflow. In accordance with the pre-registered protocol[8], two analyses were conducted. The first analysis sought to identify variables that impact the efficacy of any form of cognitive training (Aim 1: comparing training vs. control), whereas the second analysis was aimed at identifying variables that influence the efficacy of specific types of cognitive training (Aim 2: comparing different training type groups). A random forest model[80] (RF) was applied for each of these analyses, alongside a matched LME model for Aim 1.

The LME is useful for nesting participants within studies and identifying potential group moderators of outcome (i.e., training vs. control in the

Fig. 9 | Summary of analysis workflow. A visual summary of the analysis workflow.
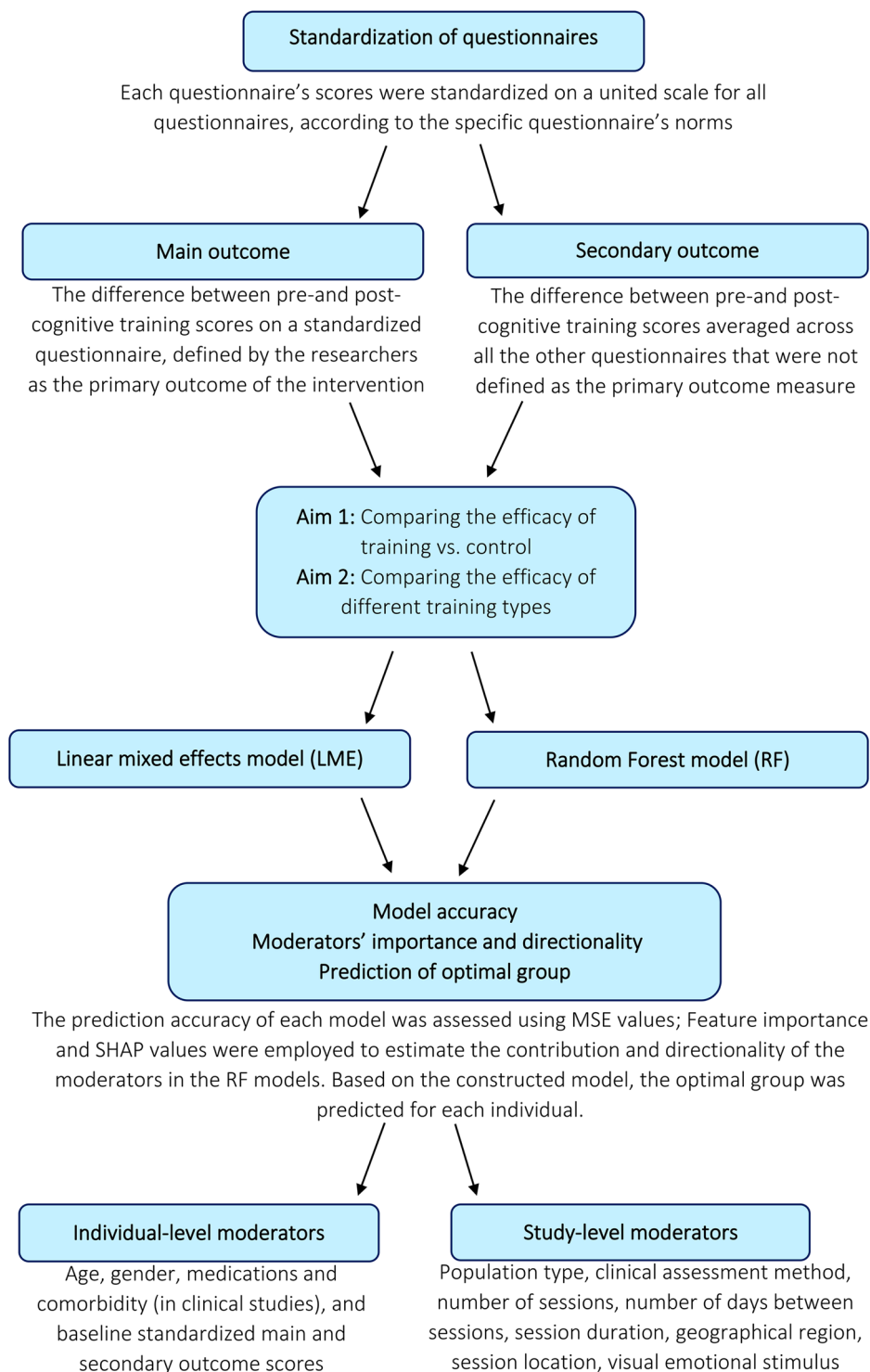
first analysis; different training types in the second analysis) by investigating the interactions between training groups, study-level moderators, and individual-level moderators. However, its primary limitation lies in its assumption of linear relationships and relative simplicity in capturing interaction effects. ML models like RF offer several advantages. RF models utilize decision trees to automatically detect and incorporate interactions between predictors by splitting outcome predictions based on potential moderators. This approach allows RF to account for complex and non-linear interactions. ML models can also handle high-dimensional data and complex relationships without needing strong assumptions about the

underlying distribution of the data. Moreover, ML techniques are more robust to overfitting and can better handle situations where there are many potential moderators[81]. For Aim 1 (comparing training vs. control), both RF and LME analyses were performed. LME analysis could not be performed for Aim 2 (comparing different training type groups) due to complete multicollinearity between the training type and the study-level moderators, and therefore only RF analysis was performed for Aim 2. In addition to the primary analysis of the full dataset, a follow-up unregistered analysis was conducted to explore Aim 1 and Aim 2 specifically among participants with a clinical diagnosis. This analysis focused on data from three studies that

included detailed clinical information, including comorbidity and medication status[58,76,77]. This step was essential to examine whether these clinical factors influence the effectiveness of cognitive training, particularly for patients, who represent most mental health care consumers. Due to the small number of participants in the follow-up analysis, only the LME model was applied.

For all analyses, the dependent variable was improvement, calculated as the difference between pre- and post-cognitive training outcome scores. In line with the pre-registered protocol[8], both Aim 1 and Aim 2 were analyzed twice, first with the standardized main outcome measure and second with the aggregated and averaged secondary outcome measure. Preprocessing for analysis of the secondary outcome included only studies that contained outcome secondary measures in addition to the reported main outcome.

The models were evaluated by calculating the mean squared error (MSE) for each analysis. MSE measures the average squared difference between predicted and actual values, calculated in units of the standard deviation of the standardized outcome scale, and offers a clear indicator of model accuracy[82].

For the RF-based analysis of the complete dataset, the tuning parameters (number of trees, minimum samples split, minimum samples leaf, maximum features, maximum depth) were selected by applying an 8-fold cross-validation procedure. The complete range of tuning parameter values is as follows:

Number of trees: 5, 10, 25, 50, 100, 200, 500, 1000, 1250, 1500
Maximum depth: 2–24.
Minimum samples leaf: 2–14
Minimum samples split: 2–14
Maximum features: 1- total number of possible moderators in each analysis.

The procedure for selecting the best tuning parameters was conducted as follows:

1. Given the extensive number of potential combinations of all tuning parameters, 500 combinations were randomly selected.
2. The final tuning parameters for each of the four models (i.e., separate analyses for Aim 1 and Aim 2, each applied to both the main and secondary outcomes) were determined using an 8-fold cross-validation procedure for each of the randomly selected combinations of tuning parameters.
3. For each combination, the sample was split into eight folds, with one-eighth of the sample withheld. The model was trained on the remaining seven folds and tested on the withheld fold to predict improvement. The mean square error (MSE) of the prediction was then calculated. Each combination received eight MSE values in this process.
4. The eight MSE values for each combination were averaged, resulting in a single cross-validated MSE (CV MSE) per combination. The combination with the lowest CV MSE was selected for each model.

For each model, the sample was randomly split into a training set (70% of the data) and a validation set (30% of the data), and the final prediction was conducted on the validation set with the optimal tuning parameters. Feature importance and Shapley Additive Explanations (SHAP) values were calculated for each model. Feature importance indicates the relative contribution of each moderator to the prediction[83], highlighting the individual and study-level characteristics that impact training efficacy. SHAP values provide insight into how and in what direction these moderators influence predictions, enhancing model interpretability and transparency. Features with positive SHAP values contribute positively to the outcome, while those with negative SHAP values have a negative effect[84].

Additionally, to determine the optimal training group for each participant, we calculated the expected gain (i.e., predicted improvement score) for both scenarios: being assigned to either the cognitive training or control group for Aim 1, and to one of the training groups for Aim 2 (i.e., working memory training, ABM, CBM-I, or inhibitory control training). Selecting the optimal group was done by using the RF model already trained and

tested on the full data set. For each participant, the expected gain was predicted separately for each optional group assignment, and the group with the highest predicted gain was selected.

For the LME analysis of the complete data set, we constructed two models, to identify the model with the best predictive performance. The first was a minimal model that included only the fixed effect of training type (training vs. control for Aim 1, and training group type for Aim 2) and the random effect of the specific study (random slope). The second was a maximal model, incorporating three-way interactions between training type, study-level moderators, and individual-level moderators, along with the random effect of the specific study.

Equation (1) is the equation for the minimal model:

$$Gain_{il} = \beta_0 + \beta \cdot TrainingGroup + b_l \cdot Study_l + \epsilon_i$$

Equation (2) is the equation for the maximal model:

$$Gain_{il} = \beta_0 + \beta \cdot TrainingGroup + \sum_k \sum_j \beta_{kj} TrainingGroup \cdot SLM_k$$
$$\cdot ILM_j + b_l \cdot Study_l + \epsilon_i$$

Where "gain" is the difference between pre-and post-training outcome scores, "i" is an observation in study "l", the sigma represents all the possible three-way interactions, and SLM and ILM represent study-level moderators and individual-level moderators, accordingly.

After constructing these models, a backward-stepwise regression using Akaike Information Criterion (AIC) was conducted to detect the optimal sub-model between the maximal and minimal models (i.e., the model with the highest number of variables included without leading to overfitting). This regression uses a data-driven stepwise selection process whereby variables are iteratively removed based on AIC, a metric that balances model fit and complexity by penalizing models with more variables to avoid overfitting[85]. This regression method was chosen because it closely resembles the process used in the RF analysis, where all variables are initially included and interactions are determined based on the data. Both approaches focus on data-driven variable selection and interaction discovery, making them similar in their handling of model complexity and feature importance. After identifying the optimal model, it was trained on 70% of the data and tested on the remaining 30% withheld validation set, following a procedure similar to that used in the RF analysis. See Supplementary Tables 1, 2, and 3 for the results of the LME analysis of the complete data set.

For the clinical data set, LME analysis was the only feasible option due to the lower number of participants. A different approach from the previous LME analysis of the full data set was necessary to prevent multicollinearity caused by the inclusion of too many moderators and interactions. The moderators included in this analysis were age, gender, medication status, type of diagnosis, training group (training vs. control for Aim 1; type of training for Aim 2), number of sessions, and baseline score (for the main or secondary outcome, in the main or secondary outcome analyses, respectively). The raw data on which this analysis was conducted included participants who took part in ABM (training or control), CBM (training or control), or a passive control group. All participants had a diagnosis of depression, with or without comorbidity.

First, a model with all moderators was constructed and then estimated. Afterward, all non-significant interactions were removed, and the model was estimated again with only the significant interactions. Multiple comparisons were subsequently performed to explore the effects further, using a Bonferroni correction and a significance threshold of $p < 0.05$.

The RF-based models were constructed using Python software version 3.8 The LME models were constructed using R software version 4.2.2.

Data availability

The data that support the findings of this study are not yet publicly available due to the large number of data contributors, but are available from the corresponding author upon request.

Code availability

The underlying code for this study is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

References

1. Nguyen, L., Murphy, K. & Andrews, G. Immediate and long-term efficacy of executive functions cognitive training in older adults: a systematic review and meta-analysis. *Psychol. Bull.* **145**, 698 (2019).
2. Siegle, G. J., Ghinassi, F. & Thase, M. E. Neurobehavioral therapies in the 21st century: summary of an emerging field and an extended example of cognitive control training for depression. *Cognit. Ther. Res.* **31**, 235–262 (2007).
3. Koster, E. H., Hoorelbeke, K., Onraedt, T., Owens, M. & Derakshan, N. Cognitive control interventions for depression: a systematic review of findings from training studies. *Clin. Psychol. Rev.* **53**, 79–92 (2017).
4. Fodor, L. A. et al. Efficacy of cognitive bias modification interventions in anxiety and depressive disorders: a systematic review and network meta-analysis. *Lancet Psychiatry* **7**, 506–514 (2020).
5. Schwaighofer, M., Fischer, F. & Bühner, M. Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educ. Psychol.* **50**, 138–166 (2015).
6. Martinelli, A., Grüll, J. & Baum, C. Attention and interpretation cognitive bias change: a systematic review and meta-analysis of bias modification paradigms. *Behav. Res. Ther.* **157**, 104180 (2022).
7. Price, R. B., Paul, B., Schneider, W. & Siegle, G. J. Neural correlates of three neurocognitive intervention strategies: a preliminary step towards personalized treatment for psychological disorders. *Cognit. Ther. Res.* **37**, 657–672 (2013).
8. Shani, R. et al. Personalized cognitive training: Protocol for individual-level meta-analysis implementing machine learning methods. *J. Psychiatr. Res.* **138**, 342–348 (2021).
9. Institute of Health Metrics and Evaluation. *Global Health Data Exchange (GHDx)*, https://vizhub.healthdata.org/gbd-results/ (2022).
10. World Health Organization, *Mental Health and COVID-19: Early evidence of the pandemic's impact*, https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci_Brief-Mental_health-2022.1 (2022).
11. Cassano, G. B., Rossi, N. B. & Pini, S. Psychopharmacology of anxiety disorders. *Dialogues Clin. Neurosci.* **4**, 271–285 (2002).
12. Ormel, J., Hollon, S. D., Kessler, R. C., Cuijpers, P. & Monroe, S. M. More treatment but no less depression: The treatment-prevalence paradox. *Clin. Psychol. Rev.* **91**, 102111 (2022).
13. Kaiser, T. et al. Heterogeneity of treatment effects in trials on psychotherapy of depression. *Clin. Psychol. Sci. Pract.* **29**, 294 (2022).
14. Green, C. S. et al. Improving methodological standards in behavioral interventions for cognitive enhancement. *J. Cognit. Enhanc.* **3**, 2–29 (2019).
15. Katz, B., Shah, P. & Meyer, D. E. How to play 20 questions with nature and lose: Reflections on 100 years of brain-training research. *Proc. Natl. Acad. Sci. USA* **115**, 9897–9904 (2018).
16. Shani, R., Tal, S., Zilcha-Mano, S. & Okon-Singer, H. Can machine learning approaches lead toward personalized cognitive training? *Front. Behav. Neurosci.* **13**, 64 (2019).
17. Linetzky, M., Pergamin-Hight, L., Pine, D. S. & Bar-Haim, Y. Quantitative evaluation of the clinical efficacy of attention bias modification treatment for anxiety disorders. *Depress. Anxiety* **32**, 383–391 (2015).
18. Price, R. B. et al. Pooled patient-level meta-analysis of children and adults completing a computer-based anxiety intervention targeting attentional bias. *Clin. Psychol. Rev.* **50**, 37–49 (2016).
19. Karbach, J., Könen, T. & Spengler, M. Who benefits the most? Individual differences in the transfer of executive control training across the lifespan. *J. Cognit. Enhanc.* **1**, 394–405 (2017).
20. Karbach, J. & Verhaeghen, P. Making working memory work: a meta-analysis of executive-control and working memory training in older adults. *Psychol. Sci.* **25**, 2027–2037 (2014).
21. Bryan, C. J., Tipton, E. & Yeager, D. S. Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nat. Hum. Behav.* **5**, 980–989 (2021).
22. Li, J., Ma, H., Yang, H., Yu, H. & Zhang, N. Cognitive bias modification for adult's depression: a systematic review and meta-analysis. *Front. Psychol.* **13**, 968638 (2023).
23. Vander Zwalmen, Y. et al. Treatment response following adaptive PASAT training for depression vulnerability: a systematic review and meta-analysis. *Neuropsychol. Rev.* **34**, 232–249 (2024).
24. Traut, H. J., Guild, R. M. & Munakata, Y. Why does cognitive training yield inconsistent benefits? A meta-analysis of individual differences in baseline cognitive abilities and training outcomes. *Front. Psychol.* **12**, 662139 (2021).
25. Bica, I., Alaa, A. M., Lambert, C. & Van Der Schaar, M. From real-world patient data to individualized treatment effects using machine learning: current and future methods to address underlying challenges. *Clin. Pharmacol. Ther.* **109**, 87–100 (2021).
26. Curth, A., Peck, R. W., McKinney, E., Weatherall, J. & van Der Schaar, M. Using machine learning to individualize treatment effect estimation: challenges and opportunities. *Clin. Pharmacol. Ther.* **115**, 710–719 (2024).
27. Vieira, S., Liang, X., Guiomar, R. & Mechelli, A. Can we predict who will benefit from cognitive-behavioural therapy? A systematic review and meta-analysis of machine learning studies. *Clin. Psychol. Rev.* **97**, 102193 (2022).
28. Feuerriegel, S. et al. Causal machine learning for predicting treatment outcomes. *Nat. Med.* **30**, 958–968 (2024).
29. Cohen, Z. D. & DeRubeis, R. J. Treatment selection in depression. *Annu. Rev. Clin. Psychol.* **14**, 209–236 (2018).
30. Richter, T., Fishbain, B., Markus, A., Richter-Levin, G. & Okon-Singer, H. Using machine learning-based analysis for behavioral differentiation between anxiety and depression. *Sci. Rep.* **10**, 1–12 (2020).
31. Richter, T., Fishbain, B., Fruchter, E., Richter-Levin, G. & Okon-Singer, H. Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders. *J. Psychiatr. Res.* **141**, 199–205 (2021).
32. Zilcha-Mano, S. Major developments in methods addressing for whom psychotherapy may work and why. *Psychother. Res.* **29**, 693–708 (2019).
33. Zilcha-Mano, S. et al. Dropout in treatment for depression: translating research on prediction into individualized treatment recommendations. *J. Clin. Psychiatry* **77**, 1584–1590 (2016).
34. Zilcha-Mano, S., Roose, S. P., Brown, P. J. & Rutherford, B. R. A machine learning approach to identifying placebo responders in late-life depression trials. *Am. J. Geriatr. Psychiatry* **26**, 669–677 (2018).
35. Jaeggi, S. M., Karbach, J. & Strobach, T. Editorial special topic: enhancing brain and cognition through cognitive training. *J. Cognit. Enhanc.* **1**, 353–357 (2017).
36. Cohen, N. & Ochsner, K. N. From surviving to thriving in the face of threats: the emerging science of emotion regulation training. *Curr. Opin. Behav. Sci.* **24**, 143–155 (2018).
37. Gober, C. D., Lazarov, A. & Bar-Haim, Y. From cognitive targets to symptom reduction: overview of attention and interpretation bias modification research. *BMJ Ment. Health* **24**, 42–46 (2021).
38. Motter, J. N. et al. Computerized cognitive training and functional recovery in major depressive disorder: a meta-analysis. *J. Affect. Disord.* **189**, 184–191 (2016).

39. von Bastian, C. C. et al. Mechanisms underlying training-induced cognitive change. *Nat. Rev. Psychol.* **1**, 30–41 (2022).

40. Lampit, A., Hallock, H. & Valenzuela, M. Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. *PLoS Med* **11**, e1001756 (2014).

41. Weicker, J., Villringer, A. & Thöne-Otto, A. Can impaired working memory functioning be improved by training? A meta-analysis with a special focus on brain-injured patients. *Neuropsychology* **30**, 190–212 (2016).

42. Yeo, H., Yoon, S., Lee, J., Kurtz, M. M. & Choi, K. A meta-analysis of the effects of social-cognitive training in schizophrenia: the role of treatment characteristics and study quality. *Br. J. Clin. Psychol.* **61**, 37–57 (2022).

43. Lejeune, J. A., Northrop, A. & Kurtz, M. M. A meta-analysis of cognitive remediation for schizophrenia: efficacy and the role of participant and treatment factors. *Schizophr. Bull.* **47**, 997–1006 (2021).

44. Vita, A. et al. Durability of effects of cognitive remediation on cognition and psychosocial functioning in schizophrenia: a systematic review and meta-analysis of randomized clinical trials. *Am. J. Psychiatry* **181**, 520–531 (2024).

45. Vita, A. et al. Effectiveness, core elements, and moderators of response of cognitive remediation for schizophrenia: a systematic review and meta-analysis of randomized clinical trials. *JAMA Psychiatry* **78**, 848–858 (2021).

46. Penney, D. et al. Immediate and sustained outcomes and moderators associated with metacognitive training for psychosis: a systematic review and meta-analysis. *JAMA Psychiatry* **79**, 417–429 (2022).

47. Edwards, E. J. et al. Cognitive control training for children with anxiety and depression: a systematic review. *J. Affect. Disord.* **300**, 158–171 (2022).

48. Jones, E. B. & Sharpe, L. Cognitive bias modification: a review of meta-analyses. *J. Affect. Disord.* **223**, 175–183 (2017).

49. Salemink, E., Woud, M. L., Bouwman, V. & Mobach, L. Cognitive bias modification training to change interpretation biases. In *Interpretational Processing Biases in Emotional Psychopathology: From Experimental Investigation to Clinical Practice* 205–226 (Springer International Publishing, 2023).

50. Richter, T., Stahi, S., Mirovsky, G., Hel-Or, H. & Okon-Singer, H. Disorder-specific versus transdiagnostic cognitive mechanisms in anxiety and depression: Machine-learning-based prediction of symptom severity. *J. Affect. Disord.* **354**, 473–482 (2024).

51. Schizophrenia International Research Society (SIRS). Research Harmonization Award. Schizophrenia Research Society https://schizophreniaresearchsociety.org/research-harmonisation-award (accessed 2024).

52. Wellcome Trust. Common metrics in mental health research. Wellcome https://wellcome.org/grant-funding/guidance/common-metrics-mental-health-research (accessed 2024).

53. International Alliance of Mental Health Research Funders (IAMHRF). Driving the adaptation of common measures. IAMHRF https://iamhrf.org/projects/driving-adoption-common-measures (accessed 2024).

54. Perlman, K. et al. Development of a differential treatment selection model for depression on consolidated and transformed clinical trial datasets. *Transl. Psychiatry* **14**, 263 (2024).

55. Chen, Z. S. et al. Modern views of machine learning for precision psychiatry. *Patterns* **3**, 100602 (2022).

56. Hegyi, P. et al. Academia Europaea position paper on translational medicine: the cycle model for translating scientific results into community benefits. *J. Clin. Med.* **9**, 1532 (2020).

57. Roessner, V. et al. Taming the chaos?! Using eXplainable Artificial Intelligence (XAI) to tackle the complexity in mental health research. *Eur. Child Adolesc. Psychiatry* **30**, 1143–1146 (2021).

58. Beevers, C. G., Clasen, P. C., Enock, P. M. & Schnyer, D. M. Attention bias modification for major depressive disorder: effects on attention bias, resting state connectivity, and symptom change. *J. Abnorm. Psychol.* **124**, 463 (2015).

59. Bunnell, B. E., Beidel, D. C. & Mesa, F. A randomized trial of attention training for generalized social phobia: does attention training change social behavior? *Behav. Ther.* **44**, 662–673 (2013).

60. Clerkin, E. M., Beard, C., Fisher, C. R. & Schofield, C. A. An attempt to target anxiety sensitivity via cognitive bias modification. *PLoS ONE* **10**, e0114578 (2015).

61. Cohen, N., Mor, N. & Henik, A. Linking executive control and emotional response: a training procedure to reduce rumination. *Clin. Psychol. Sci.* **3**, 15–25 (2015).

62. Cohen, N. & Mor, N. Enhancing reappraisal by linking cognitive control and emotion. *Clin. Psychol. Sci.* **6**, 155–163 (2018).

63. Course-Choi, J., Saville, H. & Derakshan, N. The effects of adaptive working memory training and mindfulness meditation training on processing efficiency and worry in high worriers. *Behav. Res. Ther.* **89**, 1–13 (2017).

64. Daches, S. & Mor, N. Training ruminators to inhibit negative information: a preliminary report. *Cogn. Ther. Res.* **38**, 160–171 (2014).

65. Daches, S., Mor, N. & Hertel, P. Rumination: cognitive consequences of training to inhibit the negative. *J. Behav. Ther. Exp. Psychiatry* **49**, 76–83 (2015).

66. Ducrocq, E., Wilson, M., Vine, S. & Derakshan, N. Training attentional control improves cognitive and motor task performance. *J. Sport Exerc. Psychol.* **38**, 521–533 (2016).

67. Ducrocq, E., Wilson, M., Smith, T. J. & Derakshan, N. Adaptive working memory training reduces the negative impact of anxiety on competitive motor performance. *J. Sport Exerc. Psychol.* **39**, 412–422 (2017).

68. Enock, P. M., Hofmann, S. G. & McNally, R. J. Attention bias modification training via smartphone to reduce social anxiety: a randomized, controlled multi-session experiment. *Cogn. Ther. Res.* **38**, 200–216 (2014).

69. Hotton, M., Derakshan, N. & Fox, E. A randomised controlled trial investigating the benefits of adaptive working memory training for working memory capacity and attentional control in high worriers. *Behav. Res. Ther.* **100**, 67–77 (2018).

70. Kuckertz, J. M. et al. Moderation and mediation of the effect of attention training in social anxiety disorder. *Behav. Res. Ther.* **53**, 30–40 (2014).

71. Lee, J. S. et al. How can we enhance cognitive bias modification techniques? The effects of prospective cognition. *J. Behav. Ther. Exp. Psychiatry* **49**, 120–127 (2015).

72. McNally, R. J., Enock, P. M., Tsai, C. & Tousian, M. Attention bias modification for reducing speech anxiety. *Behav. Res. Ther.* **51**, 882–888 (2013).

73. Owens, M., Koster, E. H. W. & Derakshan, N. Improving attention control in dysphoria through cognitive training: transfer effects on working memory capacity and filtering efficiency. *Psychophysiology* **50**, 297–307 (2013).

74. Rohrbacher, H. et al. Optimizing the ingredients for imagery-based interpretation bias modification for depressed mood: is self-generation more effective than imagination alone? *J. Affect. Disord.* **152**, 212–222 (2014).

75. Sari, B. A., Koster, E. H. W., Pourtois, G. & Derakshan, N. Training working memory to improve attentional control in anxiety: a proof-of-principle study using behavioral and electrophysiological measures. *Biol. Psychol.* **121**, 203–212 (2016).

76. Williams, A. D. et al. Combining imagination and reason in the treatment of depression: a randomized controlled trial of internet-based cognitive-bias modification and internet-CBT for depression. *J. Consult. Clin. Psychol.* **81**, 793–805 (2013).

77. Williams, A. D. et al. Positive imagery cognitive bias modification (CBM) and internet-based cognitive behavioral therapy (iCBT): a randomized controlled trial. *J. Affect. Disord.* **178**, 131–141 (2015).

78. Yang, W. et al. Attention bias modification training in individuals with depressive symptoms: a randomized controlled trial. *J. Behav. Ther. Exp. Psychiatry* **49**, 101–111 (2015).

79. Simons, D. J. et al. Do "brain-training" programs work? *Psychol. Sci. Public Interest* **17**, 103–186 (2016).
80. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
81. Bennett, M., Kleczyk, E. J., Hayes, K. & Mehta, R. Evaluating similarities and differences between machine learning and traditional statistical modeling in healthcare analytics. *Artificial Intelligence Annual Volume 2022* (IntechOpen, 2022).
82. Rainio, O., Teuho, J. & Klén, R. Evaluation metrics and statistical tests for machine learning. *Sci. Rep.* **14**, 6086 (2024).
83. Huynh-Thu, V. A., Saeys, Y., Wehenkel, L. & Geurts, P. Statistical interpretation of machine learning-based feature importance scores for biomarker discovery. *Bioinformatics* **28**, 1766–1774 (2012).
84. Nohara, Y., Matsumoto, K., Soejima, H. & Nakashima, N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput. Methods Programs Biomed.* **214**, 106584 (2022).
85. Cavanaugh, J. E. & Neath, A. A. The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Comput. Stat.* **11**, e1460 (2019).

Acknowledgements

Author contributions

T.R. redrafted the complete manuscript, performed data preprocessing, supervised data analysis and revised the first draft of the complete manuscript, R.S. performed data search and preprocessing, and drafted the introduction, method, and discussion sections. S.T. took part in data preprocessing, planning and executing analysis and drafting the results section. N.D., N.C., P.E., R.M., N.M., S.D., A.W., J.Y., P.C., J.K., W.Y., A.R., C.B., and B.B. contributed de-identified datasets, as well as substantially revised this manuscript. E.K. guided some of the decisions taken while planning this study design, as well as revised the protocol and the manuscript. S.Z.M. took part in planning the study, supervising, brainstorming, and reviewing. H.O.K. served as the principal supervisor for the project, took part in planning the study, design, oversight, review, and strategic development.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41746-025-01449-w.

Correspondence and requests for materials should be addressed to Thalia Richter.

Reprints and permissions information is available at http://www.nature.com/reprints

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

[1]School of Psychological Sciences, University of Haifa, Haifa, Israel. [2]The Integrated Brain and Behavior Research Center (IBBR), University of Haifa, Haifa, Israel. [3]Data Science Research Center, University of Haifa, Haifa, Israel. [4]Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. [5]Centre for Resilience and Posttraumatic Growth, National Centre for Integrative Oncology (NCIO), Reading, UK. [6]Department of Special Education, University of Haifa, Haifa, Israel. [7]The Edmond J. Safra Brain Research Center for the Study of Learning Disabilities, University of Haifa, Haifa, Israel. [8]Department of Psychology, Harvard University, Cambridge, MA, USA. [9]Department of Psychology, The Hebrew University of Jerusalem, Jerusalem, Israel. [10]Seymour Fox School of Education, The Hebrew University of Jerusalem, Jerusalem, Israel. [11]Psychology Department, Bar Ilan University, Ramat Gan, Israel. [12]School of Psychiatry, UNSW Medicine, University of New South Wales, Sydney, NSW, Australia. [13]King's College London, London, UK. [14]Department of Psychology, Stockholm University, Stockholm, Sweden. [15]School of Psychology, Korea University, Seoul, South Korea. [16]Department of Psychiatry, McLean Hospital, Belmont, MA, USA. [17]Department of Psychiatry, Harvard Medical School, Boston, MA, USA. [18]Department of Psychology, San Diego State University, San Diego, CA, USA. [19]Department of Psychology, Hunan Normal University, Hunan, China. [20]Department of Psychiatry, University of Oxford, Oxford, UK. [21]Institute for Mental Health Research and Department of Psychology, University of Texas at Austin, Austin, TX, USA. [22]Department of Psychiatry and Behavioral Neurosciences, Morsani College of Medicine, University of South Florida, Tampa, FL, USA. [23]Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium. [24]These authors contributed equally: Thalia Richter, Reut Shani. ✉e-mail: trichter@campus.haifa.ac.il