Journal of Affective Disorders 340 (2023) 213–220

Contents lists available at ScienceDirect

Journal of Affective Disorders

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

Research paper

Using machine learning to forecast symptom changes among subclinical depression patients receiving stepped care or usual care

Bruno T. Scodari [a,*], Sarah Chacko [b], Rina Matsumura [b], Nicholas C. Jacobson [a,b,c,d]

[a] Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA
[b] Center for Technology and Behavioral Health, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA
[c] Department of Psychiatry, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA
[d] Department of Computer Science, Dartmouth College, Hanover, NH, USA

ARTICLE INFO

Keywords:
Subclinical depression
Stepped care
Machine learning
PHQ-9
HADS-A

ABSTRACT

Background: Subclinical depression (SD) is a mental health disorder characterized by minor depressive symptoms. Most SD patients are treated in the primary practice, but many respond poorly to treatment at the expense of provider resources. Stepped care approaches are appealing for tiering SD care to efficiently allocate scarce resources while jointly optimizing patient outcomes. However, stepped care can be time inefficient, as some persons may respond poorly and be forced to suffer with their symptoms for prolonged periods. Machine learning can offer insight into optimal treatment paths and inform clinical recommendations for incident patients.
Methods: As part of the Step-Dep trial, participants with SD were randomized to receive stepped care (N=96) or usual care (N=140). Machine learning was used to predict changes in depressive symptoms every three months over a year for each treatment group.
Results: Tree-based models were effective in predicting PHQ-9 changes among patients who received stepped care (r=0.35–0.46, MAE=0.14–0.17) and usual care (r=0.34–0.49, MAE=0.15–0.18). Patients who received stepped care were more likely to reduce PHQ-9 scores if they had high PHQ-9 but low HADS-A scores at baseline, a low number of chronic illnesses, and an internal locus of control.
Limitations: Models may suffer from potential overfitting due to sample size limitations.
Conclusion: Our findings demonstrate the promise of machine learning for predicting changes in depressive symptoms for SD patients receiving different treatments. Trained models can intake incident patient information and predict outcomes to inform personalized care.

1. Introduction

Subclinical depression (SD) is a mental health disorder similar to major depression but characterized by reduced symptoms, shorter episodes, less comorbidities, fewer recurrences, and subdued psychosocial and physical impairments (Ayuso-Mateos et al., 2010; Kessler et al., 1997; Lyness, 2020; Wells et al., 1989). It is the strongest predictor for the onset of major depressive disorder (MDD) (Cuijpers et al., 2008; Davidson et al., 2015) and a disabling and burdensome condition itself (Beck and Koenig, 1996), affecting 9–10% of Americans over their lifetimes (Kessler et al., 1997; Lyness et al., 1999; Pietrzak et al., 2013) and contributing billions in annual economic losses to the U.S. economy (Cuijpers et al., 2007). SD is often comorbid with other medical conditions like coronary heart disease (CHD) (Schleifer et al., 1989) and/or type 2 diabetes (T2D) (Ali et al., 2006), which can lead to a perpetual cycle with each being a risk factor for the other (Katon, 2003). Such comorbidities can compound depressive symptoms and interfere with patient treatment adherence, quality of life, and survival outlook (Ali et al., 2010; Gehi et al., 2005; Katon et al., 2010; Lin et al., 2004; Sullivan et al., 2012).

The majority of SD patients who seek care are treated in the general practice (Pincus et al., 1999; Williams et al., 1995) and most commonly prescribed antidepressant medications (Lyness, 2020; Zhang et al., 2014). Unfortunately, the efficacy of antidepressants is inconsistent (Lyness, 2020) and can be ineffective among patients (Barbui et al., 2011; Baumeister, 2012; Cameron et al., 2014; Hegerl et al., 2012; Rapaport et al., 2011), enabling their symptom severity to prevail while concurrently sinking provider resources that could be better utilized in

* Corresponding author at: 1 Medical Center Drive, Lebanon, NH 03766, USA.
E-mail address: Bruno.Scodari.GR@dartmouth.edu (B.T. Scodari).

https://doi.org/10.1016/j.jad.2023.08.004
Received 5 August 2022; Received in revised form 8 March 2023; Accepted 1 August 2023
Available online 2 August 2023

an already strained system (Bodenheimer and Smith, 2013; Chisholm et al., 2004). Patient burdens can be largely reduced if incident cases are identified and triaged accordingly in at-risk populations. Individuals with chronic illnesses like CHD or T2D are prime populations for targeted interventions, as about a third have SD and of those patients, 40% will develop MDD within two years (Bot et al., 2010; Hance et al., 1996; Pols et al., 2018; Thombs et al., 2006).

The stepped care model is a new and appealing approach for tiering depression care in order to efficiently allocate provider resources while jointly optimizing patient outcomes (Andrews and World Health Organization, 2006; National Collaborating Centre for Mental Health (UK), 2011). In this system, patients start with low intensity treatments such as watchful waiting, self-help, or problem-solving exercises before progressing to high intensity treatments involving cognitive behavioral therapy or medication if positive responses are lacking (Andrews et al., 2010; Bower and Gilbody, 2005; Cuijpers et al., 2010; Gellatly et al., 2007; Richards and Richardson, 2012). Benefits of this model stem from the less resource-intensive nature of earlier treatments, as these efforts typically need not be administered by specialists and have been proposed as a cost-effective solution (van Straten et al., 2015). Despite the appeal of stepped care, current evidence is mixed regarding its efficacy (Bower and Gilbody, 2005; Muñoz et al., 2010). Some studies have found stepped care to be beneficial in populations with anxiety and depression, while other research has showed stepped care to have modest or statistically insignificant effects (Meuldijk and Wuthrich, 2019; Pols et al., 2017; van Straten et al., 2015). These mixed results reflect the etiological complexity and heterogeneity of the disorder, leading to issues in understanding who may benefit from the stepped care paradigm (Athira et al., 2020; Ménard et al., 2016).

Machine learning offers a promising way to inform personalized care for SD patients considering different treatment paths. Using historical data containing labeled outcomes for patients who received different treatments, machine learning models can be trained on separate treatment groups and subsequently leveraged to predict outcomes for new patients upon hypothetical receipt of a given treatment. Previous work has applied this framework to identify optimal treatment paths for patients with various psychiatric disorders – including depression – by modeling patient outcomes upon receipt of medications, psychotherapies, digital interventions, and neurobiological treatments (Chekroud et al., 2021). The success of these studies largely hinges upon machine learning's ability to model complex non-linear relationships between a set of predictors and an outcome, which often entails detecting and aggregating small effects and interactions among the predictors (Hastie et al., 2009). This feat is difficult to achieve using classical statistical methods that make parametric assumptions about the underlying data, model strictly linear relationships, and suffer from problems of multicollinearity. While machine learning methods may increase the risk of overfitting, they are well suited for modeling the elusive signals between patient-level predictors and heterogeneous depression outcomes.

Our study interrogates whether data from the previously conducted Step-Dep trial (Pols et al., 2017) can be used to predict future changes in depressive symptoms among SD patients with comorbid CHD and/or T2D. We used machine learning methods to model each study arm (treatment group) separately. Predictors included demographic, dietary, behavioral, and exercise-related variables, as well as depressive history and additional psychiatric measures. Feature importance was explored to understand which predictors explained depression changes within each treatment group. In the present manuscript, we sought to answer the following research questions: 1) can machine learning predict changes in depression severity among SD patients using baseline patient characteristics, and if so, 2) which features are most predictive of future changes in depressive symptoms.

2. Methods

2.1. Participants

The current sample was obtained from the open access dataset associated with the Dutch Step-Dep cluster randomized trial, which was conducted between January 2013 and November 2015 (Pols et al., 2017). Two hundred thirty-six patients with SD and comorbid T2D and/or CHD from the Netherlands consented to participate in the Step-Dep trial. Participants were identified across 27 primary centers in the Netherlands through electronic patient record systems and recruited via mail by their general practitioner. Requirements for enrollment included being 18+ years of age, possessing an International Classification of Primary Care diagnosis for T2D and/or CHD, presenting a non-depressed status per the Mini International Neuropsychiatric Interview (MINI) (van Vliet and de Beurs, 2007), and falling within the SD range according to the Patient Health Questionnaire (PHQ-9) (Kroenke and Spitzer, 2002; Lamers et al., 2008). Additional requirements included omission of any cognitive impairments, psychotic conditions, terminal illnesses, personality disorders, visual impairments, current pregnancies, use of antidepressant medication, familial losses within the past six months, history of suicide, or verbal/written impairments. Subjects were randomized to either the intervention group consisting of the stepped care treatment or control group consisting of usual care (referral to a general practitioner). After informing subjects of their randomization status and accounting for dropout, 96 patients remained in the stepped care group, and 140 remained in the usual care group. Participants were asked to complete their respective treatments and were followed for one year.

2.2. Primary outcome

The outcome of interest was change in baseline PHQ-9 score over varying time intervals (0–3, 0–6, 0–9, and 0–12 months). The PHQ-9 is a common Likert-based method used to measure depressive symptoms. It includes nine items pertinent to the Diagnostic and Statistical Manual of Mental Disorders' DSM-IV and ranges from 0–27, with higher scores being indicative of more severe depression (Bell, 1994; Kroenke and Spitzer, 2002). PHQ-9 is widely used in practice and has been shown to perform well in clinical settings (Lamers et al., 2008; Meader et al., 2011).

2.3. Features

Of the patient variables collected in the Pols et al. (2017) study, 23 were included in the open access dataset, including 15 categorical variables and 8 continuous variables. Specifically, categorical variables included sex (male/female), marital status (single/married), parental birthplace (yes/no, where yes indicates both parents born in Netherlands), rural residential area (yes/no, where yes indicates living in a rural area), employment status (employed/unemployed), education level (low/average/high), excessive alcohol usage (yes/no, where yes is >10 units/week for men and >5 units/week for women), current smoking behavior (yes/no), normal exercise behavior (yes/no, where yes is 10 minutes/day and 5 days/week), ethnic origin (Dutch/non-Dutch), onset of depression (early/late, where early is before age 55), baseline dysthymia status according to the MINI (yes/no) (Sheehan et al., 1998; van Vliet and de Beurs, 2007), and presence of comorbid illness according to the Dutch Questionnaire Chronic Illnesses for T2D (yes/no), CHD (yes/no), and comorbid T2D/CHD (yes/no). Continuous variables included the number of chronic diseases, body mass index (BMI), number of historical depressive episodes according to the Diagnostic Interview Schedule (Robins, 1981), baseline locus of control (range 0–20 with higher scores indicating a more external locus of control and lower scores indicating a more internal locus of control), baseline social support (range 0–48 with higher scores indicating more

perceived social support), baseline Hospital Anxiety and Depression Scale depression (HADS-D) and anxiety (HADS-A) scores (range 0–21 with higher scores indicating higher severity), and baseline PHQ-9 score.

2.4. Data partitioning

We partitioned the Step-Dep cohort into two distinct datasets by treatment group. In doing so, we were able to (1) train separate models to predict PHQ-9 changes upon hypothetical receipt of a specific treatment, and (2) assess the patient characteristics that explained PHQ-9 changes within each treatment group. It is also important to consider each treatment group separately due to the sheer differences in treatment regimen. Treatment in the usual care group consisted of standard access to a general practioner (GP) as provided by Dutch guidelines, whereas treatment in the intervention group was modeled after the stepped care approach introduced by van't Veer-Tazelaar et al. (2009). This consisted of four 3-month long phases that increased in intensity:

1. Step 1: Watchful waiting – No active care was provided in this initial phase, as recoveries from SD are common (van't Veer-Tazelaar et al., 2009).
2. Step 2: Written self-help course – In this step, patients were administered a written guided self-help course. Patients met with practice nurses every other week to monitor progress.
3. Step 3: Problem-solving treatment (PST) – During this phase, practice nurses provided patients with a PST to treat symptoms by focusing on practical skill building (Bosmans et al., 2012; Gask, 2006). Step-Dep's version of the PST consisted of a maximum of seven sessions during the 12-week period.
4. Step 4: Referral to GP – The last step was initiated if SD was present after completion of the subsequent steps, or if a patient was diagnosed with clinical depression or expressed suicidal ideation at any time during the follow-up period.

SD status was defined as PHQ-9$\geq$6. Patients showing recurrent SD after a 3-month interval were offered (not required) the next sequential step in the Step-Dep program. A PHQ-9$<$6 at any point during follow-up initiated Step 1 of the program. If a patient showed recurrent depressive symptoms after a period of remission, the patient was offered the next subsequent step they had not yet received. Full details of the Step-Dep treatment plan are provided in the original Pols et al. (2017) manuscript.

2.5. Machine learning

For both stepped care and usual care groups, a nested framework was used to train machine learning models for predicting changes in baseline PHQ-9 across 0–3, 0–6, 0–9, 0–12 month intervals using the 23 patient features. The outer loop of the nested framework consisted of a leave-one-out train/test split procedure. In other words, all patients (observations) were apportioned to the training set with a lone exception, which was used as the test set. To deal with missingness in the data, five iterations of multiple imputation by chained equations (MICE) (Buuren and Groothuis-Oudshoorn, 2011) were applied to the train/test split, which resulted in five different train/test combinations. The MICE object used to impute each training fold was saved and reapplied to impute the corresponding testing fold. In some instances, the MICE algorithm imputed different values for missing patient covariates across time intervals corresponding to PHQ-9 measurements. Because the patient characteristics are static, they should not change over time; therefore, for each subject where this occurred, we calculated the median value across the five observations for all columns with imputations. The median is robust to binary categorical variables, as it is equivalent to selecting the mode out of the five imputations for each patient. It is also robust to multilevel categorical variables (e.g., education level), as these variables were encoded by integers in an ordinal fashion. After imputation, extreme gradient boosting (XGBoost) (Chen and Guestrin, 2016) models were fit and hypertuned on each imputed training dataset using 5-fold cross validation with macro-averaged root mean squared error (*RMSE*) to gauge model performance in the inner loop. The final tuned model was used to make predictions for each corresponding imputed test set. Predictions for each imputed test set were pooled via macro-averaging and stored in a vector before moving onto the next iteration (Fig. 1). The resulting vector from our nested framework contained one prediction corresponding to each patient, which was used to compute overall performance metrics given the known labels for each patient. We used R-squared (R^2), Pearson correlation (r), and normalized mean absolute error (*MAE*) metrics to estimate the overall performance of our machine learning approach.

2.6. Feature importance

Shapley additive explanations (SHAP values) (Lundberg and Lee, 2017) were used to analyze feature importance for the models predicting changes in baseline PHQ-9 across the 0–12 month interval. By important variables, we refer to those that significantly influenced the magnitude and direction of the PHQ-9 predictions. For each iteration in the nested framework, SHAP values were calculated for the five models trained on imputed training sets, appended together, and stored in a data structure before continuation of the next iteration. After all iterations of the nested framework were completed, the SHAP values in the resulting data structure were pooled together via macro-averaging and visualized using feature importance summary plots.

2.7. Software

All data analysis was performed using version 4.2.0 of the R statistical software (RStudio Inc., Boston, MA, USA). Specifically, we used the *dplyr* package for data wrangling and manipulation (Wickham et al., 2022), the *mice* and *NADIA* packages for imputation (Borowski and Fic, 2022; Buuren et al., 2021), the *caret* package for machine learning (Kuhn et al., 2022), and the *SHAPforxgboost* package for feature importance analysis (Liu et al., 2021).

3. Results

3.1. Patient characteristics

Of the 236 patients included in the Step-Dep trial, 96 received stepped care and 140 received usual care. Due to randomization, patient characteristics were relatively balanced across the treatment groups (Table 1). The mean age of the total sample was 67.5 years (*SD*=10.0), 107 participants were female (45%), and the mean baseline PHQ-9 score was 9.4 (*SD*=3.2). A full description of the participant baseline characteristics is included in Table 1 (Pols et al., 2017).

3.2. Intervention uptake and loss to follow-up

In the intervention group, 90 patients (94%) started the Step-Dep treatment. A majority received the first two steps of Step-Dep, which consisted of watchful waiting and subsequent guided self-help (63% and 26%, respectively). An additional 11 patients were offered guided self-help but declined. Nine patients (9%) started the PST, and 6 additional patients (6%) declined this step. Only 3 patients progressed to the final step and were referred to their respective GP. Twenty-five patients (26%) dropped out of the intervention group due to frailty (n=7), time restraints (n=2), lack of motivation (n=7), moving away (n=2), and unknown reasons (n=7) (Pols et al., 2017). Due to loss to follow-up, 7–18% of PHQ-9 scores were missing in the intervention group, and 13–21% of PHQ-9 scores were missing in the usual care group (Supplementary Table 1). These values were imputed via MICE.

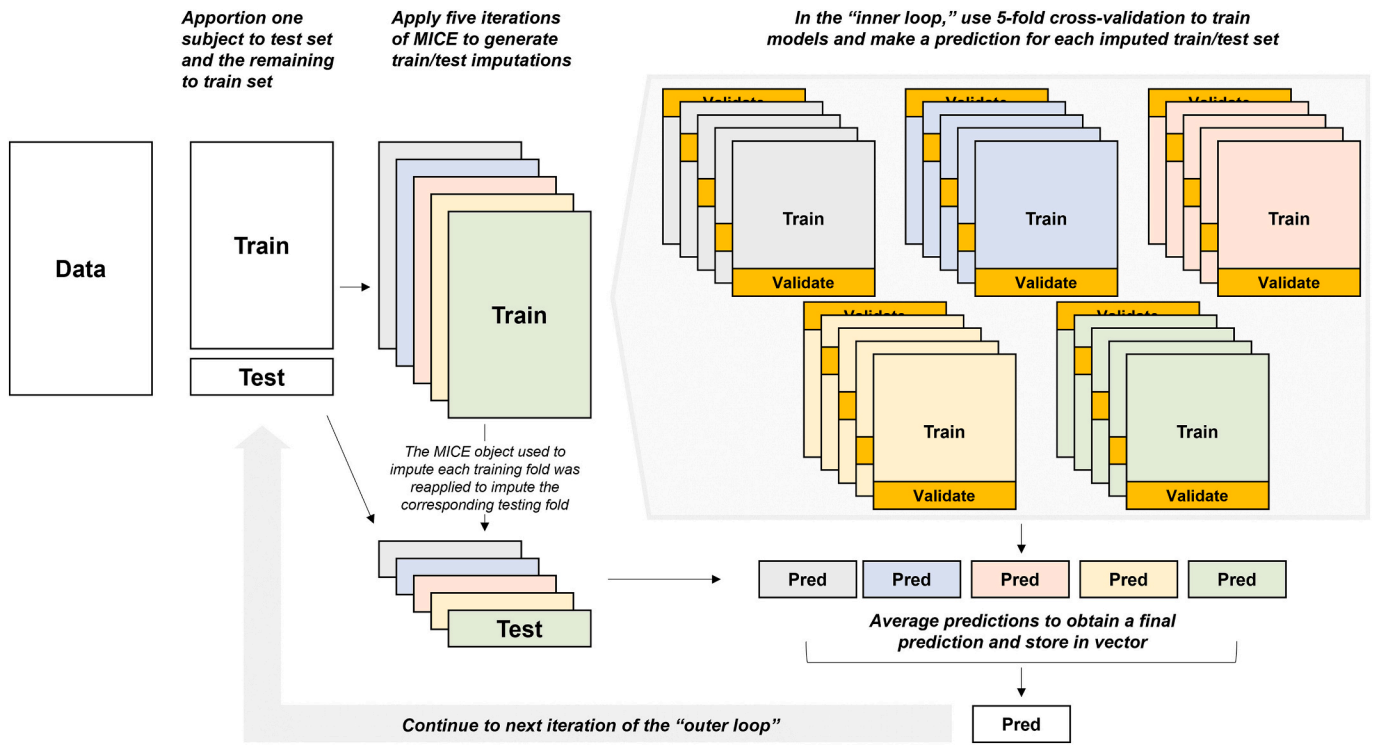

Fig. 1. Illustration of the machine learning nested framework.

3.3. Predicting PHQ-9 changes

Machine learning models were effective in predicting PHQ-9 changes for each treatment group, and varied slightly across the 0–3, 0–6, 0–9, and 0–12 month intervals (Table 2). The models trained on patients in the usual care group outperformed the models trained on patients in the intervention group for the 0–3 (r=0.49, MAE=0.16 vs. r=0.39, MAE=0.14) and 0–6 (r=0.39, MAE=0.16 vs. r=0.37, MAE=0.17) month intervals; however, the models trained on patients in the usual care group underperformed the models trained on patients in the intervention group for the 0–9 (r=0.39, MAE=0.18 vs. r=0.46, MAE=0.15) and 0–12 (r=0.34, MAE=0.15 vs. r=0.35, MAE=0.16) month intervals. Among the four models corresponding to each time interval, performance variability was similar among the usual care (SD_r=0.07, SD_{MAE}=0.01) and intervention (SD_r=0.05, SD_{MAE}=0.01) groups. Clinically significant (±5 PHQ-9 changes) improvements were predicted in 33–47% of patients in the intervention group and 29–39% of patients in the usual care group, while clinically significant regressions were predicted in 3–6% of patients in the intervention group and 4–7% of patients in the usual care group (Table 3) (Löwe et al., 2004).

3.4. Feature importance

In the models predicting PHQ-9 changes over the course of one year, SHAP plots revealed that patients who received the intervention were more likely to reduce PHQ-9 scores if they had high PHQ-9 but low HADS-A scores at baseline, a low number of chronic illnesses, and an internal locus of control (Fig. 2A). Interestingly, drivers of PHQ-9 reductions in the patients who received usual care were explained by high PHQ-9 but low HADS-D scores at baseline (Fig. 2B). The most important features varied among the intervention and usual care groups, with the only exception being baseline PHQ-9 score. Linear models were also used to generate first-order associations between the predictors and PHQ-9 changes over the course of one year for both treatment groups (Supplementary Table 2). As in the SHAP plots (Fig. 2), linear models indicated that higher PHQ-9 and lower HADS-A scores at baseline were significantly associated with depression reductions in the intervention group, while higher PHQ-9 and lower HADS-D scores at baseline were significantly associated with depression reductions in the usual care group. However, several of the meaningful drivers of PHQ-9 predictions highlighted by the SHAP plots in the intervention group were masked in the linear model.

4. Discussion

In this study, we applied machine learning methods to open access data from the Step-Dep trial to predict changes in PHQ-9 scores for patients in the intervention and usual care groups separately. We found that our models were effective in predicting the outcome in both treatment groups, and about a third of the predictions were indicative of clinically significant changes in depressive symptoms. In the intervention group, high baseline depressive symptoms (PHQ-9), low baseline anxiety symptoms (HADS-A), a low number of chronic illnesses, and an internal locus of control were the strongest predictors of reductions in depression over one year (Fig. 2A). In the usual care group, high depressive symptoms assessed by the PHQ-9 but low depressive symptoms as measured by the HADS-D at baseline were the strongest predictors of reductions in depression over one year (Fig. 2B). These models can be subsequently used to predict hypothetical PHQ-9 changes in incident SD patients and inform personalized treatment decisions concerning stepped care versus usual care.

Our results are in line with several prior studies, which also used machine learning to successfully predict similar patient outcomes among alike populations in response to antidepressants (Chekroud et al., 2016; Nie et al., 2018), psychotherapeutic techniques (Delgadillo and Gonzalez Salas Duhne, 2020), antipsychotic medications (Koutsouleris et al., 2016; Leighton et al., 2019), and emerging digital therapies (Hornstein et al., 2021; Jacobson and Nemesure, 2021). Our findings somewhat diverge from the original Pols et al. (2017) study, which concluded that Step-Dep was not more effective in preventing MDD than usual care when analyzing the same data with a linear mixed model. While this conclusion is valid given their study design and inference-

Table 1
Patient characteristics by treatment group.

	Intervention (*N*=96)	Usual care (*N*=140)	Total sample (*N*=236)
Female	42 (43.8)	65 (46.4)	107 (45.3)
Age, mean (SD)	67.8 (9.2)	67.3 (10.5)	67.5 (10.0)
Marital status			
Married/living together	55 (57.3)	67 (47.9)	122 (51.7)
Single/divorced/widowed	35 (36.5)	63 (45.0)	98 (41.5)
Not reported	6 (6.3)	10 (10.4)	16 (6.8)
Both parents born in the Netherlands	74/90 (82.2)	112/130 (86.2)	186/220 (84.5)
Rural residential area	42 (43.8)	57 (40.7)	99 (41.9)
Unemployed/sick	12/90 (13.3)	14/130 (10.8)	26/220 (11.8)
Level of education			
Low	33 (34.4)	56 (40.0)	89 (37.7)
Average	22 (22.9)	38 (27.1)	60 (25.4)
High	35 (36.5)	36 (25.7)	71 (30.1)
Not reported	6 (6.3)	10 (7.1)	16 (6.8)
Type 2 diabetes (T2D)	60 (62.5)	90 (64.3)	150 (63.6)
Coronary heart disease (CHD)	58 (60.4)	90 (64.3)	148 (62.7)
T2D and CHD	22 (22.9)	40 (28.6)	62 (26.3)
Nr of chronic diseases, median (*IQR*)	3 (2–5)	3 (2–5)	3 (2–5)
T2D treated with insulin or oral medication	42/57 (73.7)	64/83 (77.1)	106/140 (75.7)
CHD treated with chronic medication	46/54 (85.2)	65/85 (76.5)	111/139 (79.9)
Current smoker	16/90 (17.8)	23/129 (17.8)	39/219 (17.8)
Alcohol use above norm	29/90 (32.2)	34/129 (26.4)	63/219 (28.8)
Exercise under norm	56/90 (62.2)	85/129 (65.9)	141/219 (64.4)
BMI, mean (*SD*)	29.4 (6.8)	28.5 (5.6)	28.9 (6.1)
Locus of control, mean (*SD*)	8.3 (4.2)	7.6 (4.1)	7.9 (4.2)
Social support, mean (*SD*)	35.8 (9.0)	36.7 (9.5)	36.3 (9.2)
Dysthmia	6 (6.3)	7 (5.0)	13 (5.5)
Nr of depression in history			
0	35 (36.5)	65 (46.4)	100 (42.4)
1	14 (14.6)	11 (7.8)	25 (10.6)
2 or more	40 (41.7)	43 (30.7)	83 (35.2)
Not reported	7 (7.3)	21 (15.0)	28 (11.9)
Onset of depression after age of 55	38/89 (42.7)	63/121 (52.1)	101/210 (48.1)
PHQ-9 at baseline, mean (*SD*)	9.5 (3.1)	9.3 (3.2)	9.4 (3.2)
Depression HADS, mean (*SD*)	6.9 (3.9)	6.1 (3.7)	6.5 (3.8)
Anxiety HADS, mean (*SD*)	6.9 (3.7)	6.3 (3.9)	6.5 (3.8)

Note: This table was obtained from the original (Pols et al., 2017) study. Numbers in parentheses are percentages unless otherwise stated.
Abbreviations: BMI=Body Mass Index; EQ-5D-5L=Euroqol 5 dimensions 5 levels, PHQ-9=Patient Health Questionnaire-9; HADS=Hospital Anxiety and Depression Scale.

based approach, it does not recognize that some persons may respond better to stepped care than usual care. Our results are agnostic of the trial result and offer analytical evidence that individual outcomes can be accurately predicted for both treatments, enabling personalized solutions rather than generalizations for depression care.

Of particular interest are the insights gleaned from the feature importance analysis for the models predicting PHQ-9 changes over the 0–12 month interval. In the intervention group, we observed that high PHQ-9 but low HADS-A scores at baseline, a low number of chronic illnesses, and an internal locus of control were the strongest predictors of PHQ-9 reductions (Fig. 2A). Intuitively, we expected high baseline PHQ-9 scores to predict reductions in PHQ-9. The other findings also matched expectations informed by the prior literature. High anxiety levels have been shown to be associated with incident MDD in diabetic patients with SD (Bot et al., 2010), and other studies have reported that anxiety disorders are often longitudinally predictive of MDD episodes in the general population (Jacobson and Newman, 2017; King-Kallimanis et al., 2009; Merikangas et al., 2003). Similarly, the presence of chronic illness has been previously associated with incident MDD in SD populations (Cuijpers et al., 2005) as well as in elderly populations (Alvarenga et al., 2012; Niti et al., 2007). Further, prior research has consistently documented a positive relationship between external locus of control and depressive symptoms (Benassi et al., 1988; Cheng et al., 2013; Presson and Benassi, 1996). Therefore, it is unsurprising that low baseline HADS-A, a low number of chronic illnesses, and an internal locus of control were predictive of reductions in PHQ-9. On the other hand, the SHAP values for the usual care group showed that high PHQ-9 but low HADS-D scores at baseline were strong predictors of PHQ-9 reductions (Fig. 2B). This seemingly paradoxical result likely reflects known measurement variances among the PHQ-9 and HADS-D scales. The HADS-D scale focuses on measuring anhedonia symptoms (Snaith, 2003), while the PHQ-9 scale captures a more heterogeneous spectrum of depressive symptoms (Hansson et al., 2009). Thus, the usual care group's feature importance plot implies that subjects who reduced PHQ-9 scores did not experience much limitation in their perception of pleasure. Interestingly, while HADS-D was less important in explaining PHQ-9 reductions for the intervention group, the inverse was observed in the intervention group's feature importance plot despite the small dispersion of SHAP values. Considering these results altogether, it is plausible that stepped care is most effective for patients who are free of anxiety and other comorbidities, who believe they are in control of their fate, and whose perception of pleasure is limited. While this hypothesis is workable, further research is needed to explore this theory.

Our study has a few notable limitations. First, the small sample size constrains the machine learning techniques that can be explored and significantly increases the chance of overfitting. Therefore, we opted to use a nested framework and multiple imputation to hedge against overfitting while maximizing the number of observations available for

Table 2
Model performance by time interval, stratified by treatment group.

	Model 1 (*0–3 m*)	Model 2 (*0–6 m*)	Model 3 (*0–9 m*)	Model 4 (*0–12 m*)	Std. Dev. (*SD*)
Intervention (*N*=96)					
R-squared (R^2)	0.15	0.13	0.21	0.12	0.04
Pearson's correlation (*r*)	0.39	0.37	0.46	0.35	0.05
Mean absolute error (*MAE*)[a]	0.14	0.17	0.15	0.16	0.01
Usual care (*N*=140)					
R-squared (R^2)	0.24	0.15	0.15	0.11	0.05
Pearson's correlation (*r*)	0.49	0.39	0.39	0.34	0.07
Mean absolute error (*MAE*)[a]	0.16	0.16	0.18	0.15	0.01

[a] Min-max normalization used to standardize statistics to [0,1] range.

Table 3
Predictions of clinically significant[a] PHQ-9 changes by time interval, stratified by treatment group.

	Model 1 (*0–3 m*)	Model 2 (*0–6 m*)	Model 3 (*0–9 m*)	Model 4 (*0–12 m*)
Intervention (*N*=96)				
Improvement	32 (33%)	45 (47%)	39 (41%)	34 (35%)
Regression	5 (5%)	5 (5%)	3 (3%)	6 (6%)
Neither	59 (61%)	46 (48%)	54 (56%)	56 (58%)
Usual care (*N* =140)				
Improvement	44 (31%)	47 (34%)	40 (29%)	54 (39%)
Regression	10 (7%)	6 (4%)	6 (4%)	7 (5%)
Neither	86 (61%)	87 (62%)	94 (67%)	79 (56%)

[a] PHQ-9 changes of ±5 were considered clinically significant.

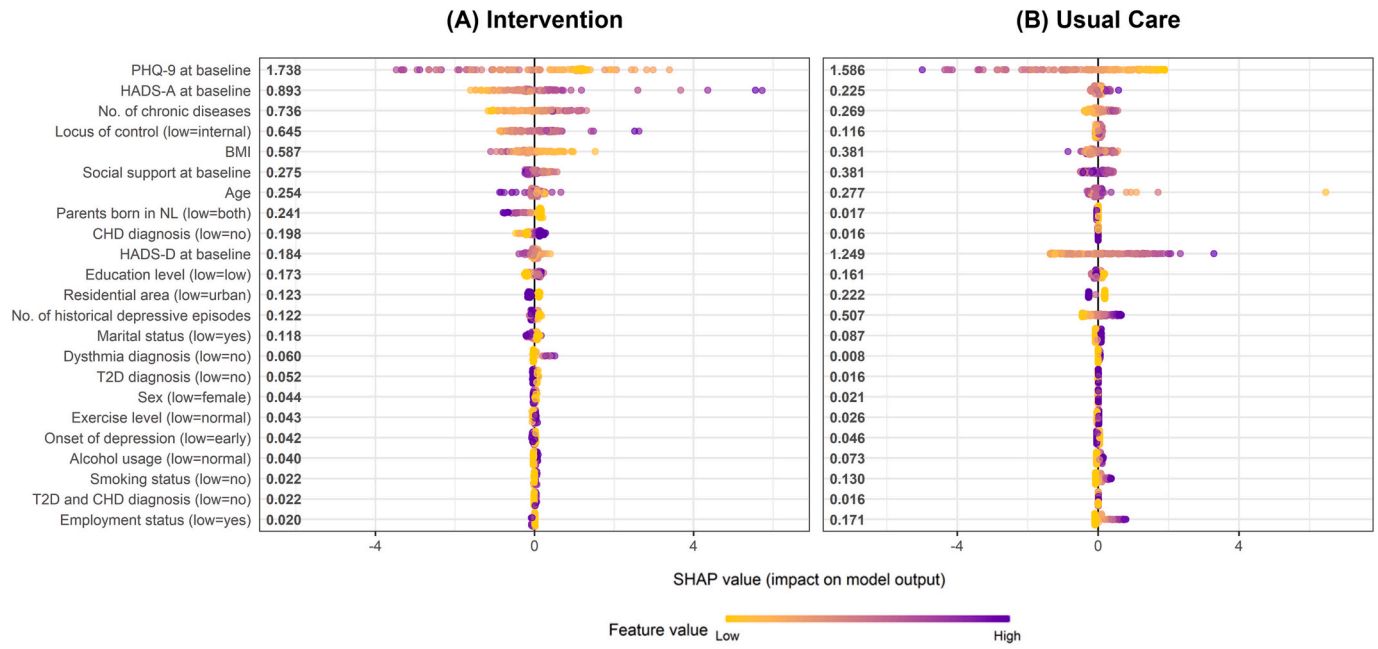

Fig. 2. SHAP feature importance plot for 12-month changes in PHQ-9 among (A) patients receiving the intervention, and (B) patients receiving usual care.

training. We recognize that many other methods could have been used with larger sample sizes that may have yielded more robust statistical properties. Second, we did not operationalize a formal rule to assign treatment modalities for incident SD patients. The naïve rule would be to assign patients to the treatment that predicts the greatest reduction (or smallest increase) in their PHQ-9 scores; however, a more clinically meaningful rule could not be established again due to the small sample size. Third, we did not attempt to develop a system to assign stepped care patients to low versus high intensity steps as done in prior research (Delgadillo et al., 2022). While we did fit a series of models to predict PHQ-9 changes across the 0–3, 0–6, 0–9, and 0–12 month intervals for the intervention group (Table 2), a very small proportion of patients randomized to stepped care progressed to high intensity treatments at later time points; therefore, creating a framework to triage patients to specific steps of the stepped care program was infeasible. Fourth, this study did not attempt to identify drivers of differential PHQ-9 predictions generated from the models trained on separate treatment groups. Future studies should seek to do so, particularly those with a larger emphasis on model inference. Lastly, we did not explore the efficacy of stepped care as a standalone intervention. Additional prospective studies should be conducted to interrogate the effectiveness of this promising intervention.

In conclusion, we found that machine learning methods were effective for predicting changes in baseline PHQ-9 scores among SD patients who received stepped care and usual care. Strong predictors of PHQ-9 reductions among stepped care patients were consistent with the primary literature and included high baseline PHQ-9, low baseline HADS-A, a low number of chronic illnesses, and an internal locus of control. Future work should seek to replicate this study with bigger data that incorporates multimodal predictors from diverse populations, pilot machine learning solutions in clinical trials, and continue identifying common psychiatric markers associated with positive stepped care outcomes. Overall, our study contributes to the mounting body of literature that uses predictive paradigms to aid in personalized depression care, contributing to the overarching goal of improving treatment while concurrently optimizing healthcare resources.

Contributors

BTS was responsible for the literature review, data wrangling & imputation, machine learning, and manuscript composition. SC and RM helped conduct the literature review and draft portions of the manuscript. Both authors made equal contributions, and the order of their names on the manuscript were determined by random simulation. NCJ conceptualized the research, provided guidance on the design and analysis plan, edited the manuscript, and supervised the study. All authors contributed to and have approved the final manuscript.

Role of the funding source

This work was partially funded by the National Institute of Mental Health (NIMH) and the National Institute of General Medical Sciences (NIGMS) [R01-MH123482-01].

Declaration of competing interest

All authors declare that they have no conflicts of interest.

Data availability

The data from the original Pols et al. (2017) study can be accessed via DataverseNL.

Acknowledgements

We would like to thank the Pols et al. (2017) research team for making the original Step-Dep dataset publicly available for further research.

Ethics approval

As the current study was based on de-identified publicly available data, it is not considered human subjects research and no ethics approval was needed for the current study. The original Step-Dep trial (NTR3715) was approved by the medical ethics committee of the VU University Medical Centre (NL39261.029.12, registration number 2012/223), and performed in accordance with the declaration of Helsinki (2008) and the Durch Medical Research involving Human Subjects Act (WMO).

Appendix A. Supplementary material

Supplementary material can be found online at https://doi.org/10.1016/j.jad.2023.08.004.

References

Ali, S., Stone, M.A., Peters, J.L., Davies, M.J., Khunti, K., 2006. The prevalence of co-morbid depression in adults with type 2 diabetes: a systematic review and meta-analysis. Diabet. Med. 23, 1165–1173. https://doi.org/10.1111/j.1464-5491.2006.01943.x.

Ali, S., Stone, M., Skinner, T.C., Robertson, N., Davies, M., Khunti, K., 2010. The association between depression and health-related quality of life in people with type 2 diabetes: a systematic literature review. Diabetes Metab. Res. Rev. 26, 75–89. https://doi.org/10.1002/dmrr.1065.

Alvarenga, M.R.M., Oliveira, M.A. de C., Faccenda, O., 2012. Sintomas depressivos em idosos: análise dos itens da Escala de Depressão Geriátrica. Acta Paul. Enferm. 25, 497–503. https://doi.org/10.1590/S0103-21002012000400003.

Andrews, G., World Health Organization (Eds.), 2006. Tolkien II: A Needs-based, Costed, Stepped-care Model for Mental Health Services: Recommendations, Executive Summaries, Clinical Pathways, Treatment Flowcharts, Costing Structures. World Health Organization, Collaborating Centre for Classification in Mental Health, Sydney, N.S.W.

Andrews, G., Cuijpers, P., Craske, M.G., McEvoy, P., Titov, N., 2010. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 5, e13196. https://doi.org/10.1371/journal.pone.0013196.

Athira, K.V., Bandopadhyay, S., Samudrala, P.K., Naidu, V.G.M., Lahkar, M., Chakravarty, S., 2020. An overview of the heterogeneity of major depressive disorder: current knowledge and future prospective. Curr. Neuropharmacol. 18, 168–187. https://doi.org/10.2174/1570159X17666191001142934.

Ayuso-Mateos, J.L., Nuevo, R., Verdes, E., Naidoo, N., Chatterji, S., 2010. From depressive symptoms to depressive disorders: the relevance of thresholds. Br. J. Psychiatry J. Ment. Sci. 196, 365–371. https://doi.org/10.1192/bjp.bp.109.071191.

Barbui, C., Cipriani, A., Patel, V., Ayuso-Mateos, J.L., van Ommeren, M., 2011. Efficacy of antidepressants and benzodiazepines in minor depression: systematic review and meta-analysis. Br. J. Psychiatry J. Ment. Sci. 198, 11–16 sup 1. https://doi.org/10.1192/bjp.bp.109.076448.

Baumeister, H., 2012. Inappropriate prescriptions of antidepressant drugs in patients with subthreshold to mild depression: time for the evidence to become practice. J. Affect. Disord. 139, 240–243. https://doi.org/10.1016/j.jad.2011.05.025.

Beck, D.A., Koenig, H.G., 1996. Minor depression: a review of the literature. Int. J. Psychiatry Med. 26, 177–209. https://doi.org/10.2190/AC30-P715-Y4TD-J7D2.

Bell, C.C., 1994. DSM-IV: diagnostic and statistical manual of mental disorders. JAMA J. Am. Med. Assoc. 272, 828. https://doi.org/10.1001/jama.1994.03520100096046.

Benassi, V.A., Sweeney, P.D., Dufour, C.L., 1988. Is there a relation between locus of control orientation and depression? J. Abnorm. Psychol. 97, 357–367. https://doi.org/10.1037//0021-843x.97.3.357.

Bodenheimer, T.S., Smith, M.D., 2013. Primary care: proposed solutions to the physician shortage without training more physicians. Health Aff. (Millwood) 32, 1881–1886. https://doi.org/10.1377/hlthaff.2013.0234.

Borowski, J., Fic, P., 2022. NADIA: NA Data Imputation Algorithms.

Bosmans, J.E., Schreuders, B., van Marwijk, H.W., Smit, J.H., van Oppen, P., van Tulder, M.W., 2012. Cost-effectiveness of problem-solving treatment in comparison with usual care for primary care patients with mental health problems: a randomized trial. BMC Fam. Pract. 13, 98. https://doi.org/10.1186/1471-2296-13-98.

Bot, M., Pouwer, F., Ormel, J., Slaets, J.P.J., de Jonge, P., 2010. Predictors of incident major depression in diabetic outpatients with subthreshold depression. Diabet. Med. J. Br. Diabet. Assoc. 27, 1295–1301. https://doi.org/10.1111/j.1464-5491.2010.03119.x.

Bower, P., Gilbody, S., 2005. Stepped care in psychological therapies: access, effectiveness and efficiency: narrative literature review. Br. J. Psychiatry 186, 11–17. https://doi.org/10.1192/bjp.186.1.11.

Buuren, S. van, Groothuis-Oudshoorn, K., 2011. mice: multivariate imputation by chained equations in R. J. Stat. Softw. 45, 1–67. https://doi.org/10.18637/jss.v045.i03.

Buuren, S. van, Groothuis-Oudshoorn, K., Vink, G., Schouten, R., Robitzsch, A., Rockenschaub, P., Doove, L., Jolani, S., Moreno-Betancur, M., White, I., Gaffert, P., Meinfelder, F., Gray, B., Arel-Bundock, V., Cai, M., Volker, T., Costantini, E., Lissa, C. van, 2021. mice: Multivariate Imputation by Chained Equations.

Cameron, I.M., Reid, I.C., MacGillivray, S.A., 2014. Efficacy and tolerability of antidepressants for sub-threshold depression and for mild major depressive disorder. J. Affect. Disord. 166, 48–58. https://doi.org/10.1016/j.jad.2014.04.078.

Chekroud, A.M., Zotti, R.J., Shehzad, Z., Gueorguieva, R., Johnson, M.K., Trivedi, M.H., Cannon, T.D., Krystal, J.H., Corlett, P.R., 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.

Chekroud, A.M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., Dwyer, D., Choi, K., 2021. The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry 20, 154–170. https://doi.org/10.1002/wps.20882.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. Association for Computing Machinery, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Cheng, C., Cheung, S.-F., Chio, J.H.-M., Chan, M.-P.S., 2013. Cultural meaning of perceived control: a meta-analysis of locus of control and psychological symptoms across 18 cultural regions. Psychol. Bull. 139, 152–188. https://doi.org/10.1037/a0028596.

Chisholm, D., Sanderson, K., Ayuso-Mateos, J.L., Saxena, S., 2004. Reducing the global burden of depression: population-level analysis of intervention cost-effectiveness in 14 world regions. Br. J. Psychiatry J. Ment. Sci. 184, 393–403. https://doi.org/10.1192/bjp.184.5.393.

Cuijpers, P., Smit, F., Willemse, G., 2005. Predicting the onset of major depression in subjects with subthreshold depression in primary care: a prospective study. Acta Psychiatr. Scand. 111, 133–138. https://doi.org/10.1111/j.1600-0447.2004.00416.x.

Cuijpers, P., Smit, F., Oostenbrink, J., De Graaf, R., Ten Have, M., Beekman, A., 2007. Economic costs of minor depression: a population-based study. Acta Psychiatr. Scand. 115, 229–236. https://doi.org/10.1111/j.1600-0447.2006.00851.x.

Cuijpers, P., van Straten, A., Smit, F., Mihalopoulos, C., Beekman, A., 2008. Preventing the onset of depressive disorders: a meta-analytic review of psychological interventions. Am. J. Psychiatry 165, 1272–1280. https://doi.org/10.1176/appi.ajp.2008.07091422.

Cuijpers, P., Donker, T., van Straten, A., Li, J., Andersson, G., 2010. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol. Med. 40, 1943–1957. https://doi.org/10.1017/S0033291710000772.

Davidson, S.K., Harris, M.G., Dowrick, C.F., Wachtler, C.A., Pirkis, J., Gunn, J.M., 2015. Mental health interventions and future major depression among primary care patients with subthreshold depression. J. Affect. Disord. 177, 65–73. https://doi.org/10.1016/j.jad.2015.02.014.

Delgadillo, J., Gonzalez Salas Duhne, P., 2020. Targeted prescription of cognitive-behavioral therapy versus person-centered counseling for depression using a machine learning approach. J. Consult. Clin. Psychol. 88, 14–24. https://doi.org/10.1037/ccp0000476.

Delgadillo, J., Ali, S., Fleck, K., Agnew, C., Southgate, A., Parkhouse, L., Cohen, Z.D., DeRubeis, R.J., Barkham, M., 2022. Stratified care vs stepped care for depression: a cluster randomized clinical trial. JAMA Psychiatry 79, 101–108. https://doi.org/10.1001/jamapsychiatry.2021.3539.

Gask, L., 2006. Problem-solving treatment for anxiety and depression: a practical guide by Laurence Mynors-Wallis. Oxford and New York: Oxford University Press. 2005. 206pp. £24.95 (pb). ISBN 0198528426. Br. J. Psychiatry 189, 287–288. https://doi.org/10.1192/bjp.189.3.287a.

Gehi, A., Haas, D., Pipkin, S., Whooley, M.A., 2005. Depression and medication adherence in outpatients with coronary heart disease: findings from the heart and soul study. Arch. Intern. Med. 165, 2508–2513. https://doi.org/10.1001/archinte.165.21.2508.

Gellatly, J., Bower, P., Hennessy, S., Richards, D., Gilbody, S., Lovell, K., 2007. What makes self-help interventions effective in the management of depressive symptoms? Meta-analysis and meta-regression. Psychol. Med. 37, 1217–1228. https://doi.org/10.1017/S0033291707000062.

Hance, M., Carney, R.M., Freedland, K.E., Skala, J., 1996. Depression in patients with coronary heart disease. A 12-month follow-up. Gen. Hosp. Psychiatry 18, 61–65. https://doi.org/10.1016/0163-8343(95)00100-x.

Hansson, M., Chotai, J., Nordström, A., Bodlund, O., 2009. Comparison of two self-rating scales to detect depression: HADS and PHQ-9. Br. J. Gen. Pract. 59, e283–e288. https://doi.org/10.3399/bjgp09X454070.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. In: Springer Series in Statistics. Springer, New York, NY. https://doi.org/10.1007/978-0-387-84858-7.

Hegerl, U., Schönknecht, P., Mergl, R., 2012. Are antidepressants useful in the treatment of minor depression: a critical update of the current literature. Curr. Opin. Psychiatry 25, 1–6. https://doi.org/10.1097/YCO.0b013e32834dc15d.

Hornstein, S., Forman-Hoffman, V., Nazander, A., Ranta, K., Hilbert, K., 2021. Predicting therapy outcome in a digital mental health intervention for depression and anxiety: a machine learning approach. Digit. Health 7, 20552076211060660. https://doi.org/10.1177/20552076211060659.

Jacobson, N.C., Nemesure, M.D., 2021. Using artificial intelligence to predict change in depression and anxiety symptoms in a digital intervention: evidence from a transdiagnostic randomized controlled trial. Psychiatry Res. 295, 113618 https://doi.org/10.1016/j.psychres.2020.113618.

Jacobson, N.C., Newman, M.G., 2017. Anxiety and depression as bidirectional risk factors for one another: a meta-analysis of longitudinal studies. Psychol. Bull. 143, 1155–1200. https://doi.org/10.1037/bul0000111.

Katon, W.J., 2003. Clinical and health services relationships between major depression, depressive symptoms, and general medical illness. Biol. Psychiatry 54, 216–226. https://doi.org/10.1016/s0006-3223(03)00273-7.

Katon, W.J., Lin, E.H.B., Von Korff, M., Ciechanowski, P., Ludman, E., Young, B., Rutter, C., Oliver, M., McGregor, M., 2010. Integrating depression and chronic disease care among patients with diabetes and/or coronary heart disease: the design of the TEAMcare study. Contemp. Clin. Trials 31, 312–322. https://doi.org/10.1016/j.cct.2010.03.009.

Kessler, R.C., Zhao, S., Blazer, D.G., Swartz, M., 1997. Prevalence, correlates, and course of minor depression and major depression in the national comorbidity survey. J. Affect. Disord. 45, 19–30. https://doi.org/10.1016/S0165-0327(97)00056-6.

King-Kallimanis, B., Gum, A.M., Kohn, R., 2009. Comorbidity of depressive and anxiety disorders for older Americans in the national comorbidity survey-replication. Am. J. Geriatr. Psychiatry 17, 782–792. https://doi.org/10.1097/JGP.0b013e3181ad4d17.

Koutsouleris, N., Kahn, R.S., Chekroud, A.M., Leucht, S., Falkai, P., Wobrock, T., Derks, E.M., Fleischhacker, W.W., Hasan, A., 2016. Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. Lancet Psychiatry 3, 935–946. https://doi.org/10.1016/S2215-0366(16)30171-7.

Kroenke, K., Spitzer, R.L., 2002. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr. Ann. 32, 509–515. https://doi.org/10.3928/0048-5713-20020901-06.

Kuhn, M. cre, Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T., 2022. caret: Classification and Regression Training.

Lamers, F., Jonkers, C.C.M., Bosma, H., Penninx, B.W.J.H., Knottnerus, J.A., van Eijk, J. Th.M., 2008. Summed score of the Patient Health Questionnaire-9 was a reliable and valid method for depression screening in chronically ill elderly patients. J. Clin. Epidemiol. 61, 679–687. https://doi.org/10.1016/j.jclinepi.2007.07.018.

Leighton, S.P., Upthegrove, R., Krishnadas, R., Benros, M.E., Broome, M.R., Gkoutos, G. V., Liddle, P.F., Singh, S.P., Everard, L., Jones, P.B., Fowler, D., Sharma, V., Freemantle, N., Christensen, R.H.B., Albert, N., Nordentoft, M., Schwannauer, M., Cavanagh, J., Gumley, A.I., Birchwood, M., Mallikarjun, P.K., 2019. Development and validation of multivariable prediction models of remission, recovery, and quality of life outcomes in people with first episode psychosis: a machine learning approach. Lancet Digit. Health 1, e261–e270. https://doi.org/10.1016/S2589-7500(19)30121-9.

Lin, E.H.B., Katon, W., Von Korff, M., Rutter, C., Simon, G.E., Oliver, M., Ciechanowski, P., Ludman, E.J., Bush, T., Young, B., 2004. Relationship of depression and diabetes self-care, medication adherence, and preventive care. Diabetes Care 27, 2154–2160. https://doi.org/10.2337/diacare.27.9.2154.

Liu, Y. cre, Just [aut, A. ctb, Mayer, M., 2021. SHAPforxgboost: SHAP Plots for "XGBoost".

Löwe, B., Unützer, J., Callahan, C.M., Perkins, A.J., Kroenke, K., 2004. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med. Care 42, 1194–1201. https://doi.org/10.1097/00005650-200412000-00006.

Lundberg, S., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. https://doi.org/10.48550/arXiv.1705.07874.

Lyness, J.M., 2020. Unipolar minor depression in adults: management [WWW Document]. URL. https://www.medilib.ir/uptodate/show/87127 (accessed 7.26.22).

Lyness, J.M., King, D.A., Cox, C., Yoediono, Z., Caine, E.D., 1999. The importance of subsyndromal depression in older primary care patients: prevalence and associated functional disability. J. Am. Geriatr. Soc. 47, 647–652. https://doi.org/10.1111/j.1532-5415.1999.tb01584.x.

Meader, N., Mitchell, A.J., Chew-Graham, C., Goldberg, D., Rizzo, M., Bird, V., Kessler, D., Packham, J., Haddad, M., Pilling, S., 2011. Case identification of depression in patients with chronic physical health problems: a diagnostic accuracy meta-analysis of 113 studies. Br. J. Gen. Pract. 61, e808–e820. https://doi.org/10.3399/bjgp11X613151.

Ménard, C., Hodes, G.E., Russo, S.J., 2016. Pathogenesis of depression: insights from human and rodent studies. Neuroscience 321, 138–162. https://doi.org/10.1016/j.neuroscience.2015.05.053.

Merikangas, K.R., Zhang, H., Avenevoli, S., Acharyya, S., Neuenschwander, M., Angst, J., Zurich Cohort Study, 2003. Longitudinal trajectories of depression and anxiety in a prospective community study: the Zurich Cohort Study. Arch. Gen. Psychiatry 60, 993–1000. https://doi.org/10.1001/archpsyc.60.9.993.

Meuldijk, D., Wuthrich, V.M., 2019. Stepped-care treatment of anxiety and depression in older adults: a narrative review. Aust. J. Rural Health 27, 275–280. https://doi.org/10.1111/ajr.12524.

Muñoz, R.F., Cuijpers, P., Smit, F., Barrera, A.Z., Leykin, Y., 2010. Prevention of major depression. Annu. Rev. Clin. Psychol. 6, 181–212. https://doi.org/10.1146/annurev-clinpsy-033109-132040.

National Collaborating Centre for Mental Health (UK), 2011. Common Mental Health Disorders: Identification and Pathways to Care. National Institute for Health and Care Excellence: Guidelines. British Psychological Society (UK), Leicester (UK).

Nie, Z., Vairavan, S., Narayan, V.A., Ye, J., Li, Q.S., 2018. Predictive modeling of treatment resistant depression using data from STAR*D and an independent clinical study. PLoS One 13, e0197268. https://doi.org/10.1371/journal.pone.0197268.

Niti, M., Ng, T.-P., Kua, E.H., Ho, R.C.M., Tan, C.H., 2007. Depression and chronic medical illnesses in Asian older adults: the role of subjective health and functional status. Int. J. Geriatr. Psychiatry 22, 1087–1094. https://doi.org/10.1002/gps.1789.

Pietrzak, R.H., Kinley, J., Afifi, T.O., Enns, M.W., Fawcett, J., Sareen, J., 2013. Subsyndromal depression in the United States: prevalence, course, and risk for incident psychiatric outcomes. Psychol. Med. 43, 1401–1414. https://doi.org/10.1017/S0033291712002309.

Pincus, H.A., Davis, W.W., McQueen, L.E., 1999. "Subthreshold" mental disorders. A review and synthesis of studies on minor depression and other "brand names". Br. J. Psychiatry J. Ment. Sci. 174, 288–296. https://doi.org/10.1192/bjp.174.4.288.

Pols, A.D., van Dijk, S.E., Bosmans, J.E., Hoekstra, T., van Marwijk, H.W.J., van Tulder, M.W., Adriaanse, M.C., 2017. Effectiveness of a stepped-care intervention to prevent major depression in patients with type 2 diabetes mellitus and/or coronary heart disease and subthreshold depression: a pragmatic cluster randomized controlled trial. PLoS One 12, e0181023. https://doi.org/10.1371/journal.pone.0181023.

Pols, A.D., Adriaanse, M.C., van Tulder, M.W., Heymans, M.W., Bosmans, J.E., van Dijk, S.E., van Marwijk, H.W.J., 2018. Two-year effectiveness of a stepped-care depression prevention intervention and predictors of incident depression in primary care patients with diabetes type 2 and/or coronary heart disease and subthreshold depression: data from the Step-Dep cluster randomised controlled trial. BMJ Open 8, e020412. https://doi.org/10.1136/bmjopen-2017-020412.

Presson, P.K., Benassi, V.A., 1996. Locus of control orientation and depressive symptomatology: a meta-analysis. J. Soc. Behav. Pers. 11, 201–212.

Rapaport, M.H., Nierenberg, A.A., Howland, R., Dording, C., Schettler, P.J., Mischoulon, D., 2011. The treatment of minor depression with St. John's Wort or citalopram: failure to show benefit over placebo. J. Psychiatr. Res. 45, 931–941. https://doi.org/10.1016/j.jpsychires.2011.05.001.

Richards, D., Richardson, T., 2012. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin. Psychol. Rev. 32, 329–342. https://doi.org/10.1016/j.cpr.2012.02.004.

Robins, L.N., 1981. National Institute of Mental Health diagnostic interview schedule: its history, characteristics, and validity. Arch. Gen. Psychiatry 38, 381. https://doi.org/10.1001/archpsyc.1981.01780290015001.

Schleifer, S.J., Macari-Hinson, M.M., Coyle, D.A., Slater, W.R., Kahn, M., Gorlin, R., Zucker, H.D., 1989. The nature and course of depression following myocardial infarction. Arch. Intern. Med. 149, 1785–1789.

Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J. Clin. Psychiatry 59 (Suppl. 20), 22–33 quiz 34-57.

Snaith, R.P., 2003. The hospital anxiety and depression scale. Health Qual. Life Outcomes 1, 29. https://doi.org/10.1186/1477-7525-1-29.

Sullivan, M.D., O'Connor, P., Feeney, P., Hire, D., Simmons, D.L., Raisch, D.W., Fine, L. J., Narayan, K.M.V., Ali, M.K., Katon, W.J., 2012. Depression predicts all-cause mortality: epidemiological evaluation from the ACCORD HRQL substudy. Diabetes Care 35, 1708–1715. https://doi.org/10.2337/dc11-1791.

Thombs, B.D., Bass, E.B., Ford, D.E., Stewart, K.J., Tsilidis, K.K., Patel, U., Fauerbach, J. A., Bush, D.E., Ziegelstein, R.C., 2006. Prevalence of depression in survivors of acute myocardial infarction. J. Gen. Intern. Med. 21, 30–38. https://doi.org/10.1111/j.1525-1497.2005.00269.x.

van Straten, A., Hill, J., Richards, D.A., Cuijpers, P., 2015. Stepped care treatment delivery for depression: a systematic review and meta-analysis. Psychol. Med. 45, 231–246. https://doi.org/10.1017/S0033291714000701.

van Vliet, I.M., de Beurs, E., 2007. The MINI-International Neuropsychiatric Interview. A brief structured diagnostic psychiatric interview for DSM-IV en ICD-10 psychiatric disorders. Tijdschr. Voor Psychiatr. 49, 393–397.

van't Veer-Tazelaar, P.J., van Marwijk, H.W.J., van Oppen, P., van Hout, H.P.J., van der Horst, H.E., Cuijpers, P., Smit, F., Beekman, A.T.F., 2009. Stepped-care prevention of anxiety and depression in late life: a randomized controlled trial. Arch. Gen. Psychiatry 66, 297. https://doi.org/10.1001/archgenpsychiatry.2008.555.

Wells, K.B., Stewart, A., Hays, R.D., Burnam, M.A., Rogers, W., Daniels, M., Berry, S., Greenfield, S., Ware, J., 1989. The functioning and well-being of depressed patients. Results from the Medical Outcomes Study. JAMA 262, 914–919.

Wickham, H., François, R., Henry, L., Müller, K., RStudio, 2022. dplyr: A Grammar of Data Manipulation.

Williams, J.W., Kerber, C.A., Mulrow, C.D., Medina, A., Aguilar, C., 1995. Depressive disorders in primary care: prevalence, functional disability, and identification. J. Gen. Intern. Med. 10, 7–12. https://doi.org/10.1007/BF02599568.

Zhang, D.X., Lewis, G., Araya, R., Tang, W.K., Mak, W.W.S., Cheung, F.M.C., Mercer, S. W., Griffiths, S.M., Woo, J., Lee, D.T.F., Kung, K., Lam, A.T., Yip, B.H.K., Wong, S.Y. S., 2014. Prevention of anxiety and depression in Chinese: a randomized clinical trial testing the effectiveness of a stepped care program in primary care. J. Affect. Disord. 169, 212–220. https://doi.org/10.1016/j.jad.2014.08.015.