Journal of Anxiety Disorders 110 (2025) 102978

Contents lists available at ScienceDirect

Journal of Anxiety Disorders

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

Development of a machine learning-based multivariable prediction model for the naturalistic course of generalized anxiety disorder

Candice Basterfield *, Michelle G. Newman

Pennsylvania State University, USA

ARTICLE INFO

Keywords:
Generalized anxiety disorder
Machine learning
Classification
Elastic net
Gradient boosted trees

ABSTRACT

Background: Generalized Anxiety Disorder (GAD) is a chronic condition. Enabling the prediction of individual trajectories would facilitate tailored management approaches for these individuals. This study used machine learning techniques to predict the recovery of GAD at a nine-year follow-up.

Method: The study involved 126 participants with GAD. Various baseline predictors from psychological, social, biological, sociodemographic and health variables were used. Two machine learning models, gradient boosted trees, and elastic nets were compared to predict the clinical course in participants with GAD.

Results: At nine-year follow-up, 95 participants (75.40 %) recovered. Elastic nets achieved a cross-validated area-under-the-receiving-operator-characteristic-curve (AUC) of .81 and a balanced accuracy of 72 % (sensitivity of .70 and specificity of .76). The elastic net algorithm revealed that the following factors were highly predictive of nonrecovery at follow-up: higher depressed affect, experiencing daily discrimination, more mental health professional visits, and more medical professional visits. The following variables predicted recovery: having some college education or higher, older age, more friend support, higher waist-to-hip ratio, and higher positive affect.

Conclusions: There was acceptable performance in predicting recovery or nonrecovery at a nine-year follow-up. This study advances research on GAD outcomes by understanding predictors associated with recovery or nonrecovery. Findings can potentially inform more targeted preventive interventions, ultimately improving care for individuals with GAD. This work is a critical first step toward developing reliable and feasible machine learning-based predictions for applications to GAD.

The trajectory of generalized anxiety disorder (GAD) is chronic and fluctuating. Epidemiological and clinical research revealed that GAD had a low probability of recovery and a high probability of recurrence (Bruce et al., 2005; Penninx et al., 2011; Scholten et al., 2013). Nearly half of all individuals with a lifetime history of the disorder still experienced it 12 months later (Martin, 2003; Ramsawh et al., 2009). Over two years, just 39 % of those initially diagnosed with GAD made a complete recovery. However, 30 % experienced a full relapse (Rodriguez et al., 2006). Up to six years after their initial anxiety disorder diagnosis, around 60 % of individuals had recurrent symptoms and relapses (Batelaan et al., 2014; Spinhoven et al., 2016). GAD course in those who relapse is often characterized by substantial levels of disability in social interactions, work engagement, and other vital aspects of life (American Psychiatric Association, 2013; Newman et al., 2022).

The chronic and fluctuating nature of GAD is influenced and exacerbated by various factors. Variables affecting its trajectory include the state of familial relationships, the presence of coexisting conditions, and gender (Rhebergen et al., 2017). Weak ties with spouses or relatives and the existence of accompanying cluster C personality disorders have been linked with a decreased probability of GAD remission (Yonkers et al., 2000). Additionally, the presence of comorbid Axis I disorders has shown an impact on its course; individuals with GAD who concurrently experienced depression, panic disorder with agoraphobia, or substance use disorders had reduced chance of recovering compared to those without these comorbidities (Bruce et al., 2005). Moreover, women had lower rates of remission than men, but showed greater stability, as evidenced by a reduced tendency for relapse (Yonkers et al., 2003).

Despite various predictors at the group level, the accuracy of the prediction of GAD outcomes for individual patients remains uncertain. No all-encompassing model is available with the necessary sensitivity and specificity for predicting its course in a manner that can be practically applied for individual use. Distinct diagnoses of anxiety disorders, according to the Diagnostic Statistical Manual (DSM; American

* Correspondence to: Department of Psychology, Pennsylvania State University, State College, PA 16803, USA
E-mail address: cbb5563@psu.edu (C. Basterfield).

https://doi.org/10.1016/j.janxdis.2025.102978
Received 9 May 2024; Received in revised form 16 January 2025; Accepted 19 January 2025
Available online 25 January 2025
0887-6185/© 2025 Elsevier Ltd. All rights reserved, including those for text and data mining, AI training, and similar technologies.

Psychiatric Association, 2013), have demonstrated limited efficacy in predicting psychopathology course (Batelaan et al., 2014). Without advanced models, GAD prediction relies on clinician judgment, which has shown poor accuracy (Bowes et al., 2020; Meehl, 1954).

A potential reason for the poor accuracy in predicting course outcomes in GAD may be attributed to the intricate and multifaceted origins of anxiety disorders (Mineka & Zinbarg, 2006; Newman et al., 2013; Penninx et al., 2008). Studies have yielded significant findings, but single risk factors only explained parts of GAD etiology (Mineka & Zinbarg, 2006; Penninx et al., 2008). Given the wide array of potential influences, it is unlikely that there exists one single predictor that explains all of the variance in the naturalistic course of GAD. Therefore, it is essential to combine predictors from psychological, social, and biological domains (Merikangas & Pine, 2002). Predictive models based on machine learning are gaining popularity as they can incorporate a large amount of data into a single model, focusing on optimizing the assessment of the model's ability to predict outcomes for individuals. Thus, machine learning may have potential not encountered before for predicting GAD trajectories (Hahn et al., 2017). Such trajectories can be regarded as a classification problem where the algorithm assigns a value to a specific class, and in the case of GAD, the prediction would be based on recovery and nonrecovery classes, which can be solved using supervised machine learning algorithms. These algorithms are trained on participants with known predictors and outcome data to establish a function capable of predicting outcomes for new participants based on their predictor values (Bokma et al., 2022).

The current study was the first one that aimed to use machine learning to predict naturalistic remission and nonremission 9 years later in individuals with a baseline diagnosis of GAD. Prior studies using machine learning in anxiety disorders have mostly focused on different research questions than the current study. For example, Li et al. (2024) used machine learning to predict naturalistic onset of any anxiety disorder in participants with no anxiety disorders at baseline. Other researchers applied machine learning models to understand the importance of clinical biomarkers for the concurrent diagnosis of anxiety disorders (Calderon et al., 2024; Sharma & Verbeke, 2021) or to predict response to specific GAD treatments (Gyorda et al., 2023; Lueken et al., 2016; Zainal & Newman, 2024; Zainal & Newman, in press). Also, in the prior studies listed above, baseline predictors were examined for more short-term outcomes (i.e., 6 days to 3 years) than the current study. Similarly, the one study that examined prediction of naturalistic remission/nonremission did so for those with any anxiety disorder at baseline and predicted outcome 2-years later Bokma et al. (2022). Thus, by examining predictors of 9-year remission/non-remission of GAD, our study might be able to identify specific baseline variables that could be targeted by GAD specific interventions to prevent longterm nonrecovery or recurrence of GAD.

We compared two machine learning models, gradient boosted trees and elastic nets, to predict recovery or nonrecovery in participants with GAD. Baseline variables examined as predictors included clinical, psychological, biological, sociodemographic, and lifestyle data from the Midlife Development in the United States (MIDUS) study. Our primary outcome was recovery or nonrecovery from GAD at a nine-year follow-up. Finally, we assessed which predictor domains contributed most to recovery or nonrecovery. We hypothesized that gradient boosted trees and elastic nets using a wide array of baseline data from different domains would yield adequate nine-year recovery predictions.

1. Method

1.1. Participants

Data in the current study were drawn from MIDUS, a national, longitudinal collection that focused on the role of psychological, social, biological, sociodemographic, and health variables. The initial MIDUS 1 probability sample (N = 7100) was generated in 1995–1996 through random digit dialing of U.S. households with at least one telephone in the contiguous 48 states, stratified by age with an oversample of those between 40 and 60. The second measurement occasion, MIDUS 2, was 9 years later, and 75 % of the original sample, adjusted for mortality (N = 4955), was retested (Radler & Ryff, 2010).

In the present study, we used data from participants who satisfied the following criteria: presence of 12-month GAD diagnosis at baseline (MIDUS 1), based on the third edition of the DSM (American Psychiatric Association, 1980), and availability of nine-year follow-up on diagnosis of GAD (recovery or nonrecovery). In our sample, psychiatric comorbidity was allowed. Participants with GAD diagnosis at baseline were selected (n = 192). Of these participants, 66 were excluded due to missing diagnostic information at a nine-year follow-up. This yielded a sample of 126 GAD participants with sufficient data available. Mean age was 43.17 (SD = 10.66), ranging between 25 and 68, and 73 % of participants identified as female. Concerning race, 87 % of participants identified as White, 6 % as Black, 2 % as Native American, and 5 % as Other. In terms of education, 49 % had graduated high school or less, and 51 % had attended college or more. The primary outcome classification task in the current machine learning study was predicting recovery or nonrecovery from GAD at a nine-year follow-up.

1.2. Baseline predictor variables

At baseline, a wide array of predictors, including clinical, psychological, sociodemographic, biological, and lifestyle, were included, making up 80 variables.

1.2.1. Sociodemographic factors

Age, gender, race/ethnicity, marital status, and education.

1.2.2. Life challenges

Daily stressors (e.g., work overload, family arguments, traffic problems), chronic stressors (e.g., caregiving, perceived discrimination, perceived inequalities, work-family spillover, childcare difficulties, unemployment), acute events (e.g., divorce, remarriage, job change, deaths, relocation).

1.2.3. Health behaviors

Smoking, alcohol consumption, physical activity, substance abuse, hormone therapy, preventive healthcare, alternative healthcare.

1.2.4. Psychological

Personality, affect, coping, control, goal orientation, optimism, religion/spirituality, and health beliefs.

1.2.5. Social

Social support, spousal relations, parent-child ties, childhood violence, social participation, social responsibility, job characteristics, and neighborhood quality.

1.2.6. Health/Illness

Mental (depression, anxiety, psychological well-being, and cognitive function); physical (subjective health, health comparisons, chronic conditions, symptoms, and disability/functional limitations).

1.2.7. Baseline comorbid conditions

A count of the number of comorbid conditions.

1.3. Statistical analyses and machine learning algorithms

1.3.1. Data preprocessing

For this study, we followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis guidelines (TRIPOD; Moons et al., 2015). There was no missing data in the current analyses. According to Occam's Razor, also known as the Law of

Parsimony, the best explanation of a problem requires the fewest assumptions. In line with Occam's Razor, the primary goal of machine learning is to enhance the efficiency of predictive models. Initially, there were 80 potential predictors. However, considering our sample size ($N = 126$), we followed established heuristics that recommend a minimum of 12 observations per candidate predictor (Jenkins & Quintana-Ascencio, 2020). Thus, we initially ran a logistic regression with all 80 potential predictors and then selected the top 11 with the strongest coefficients. Logistic regression is a widely used classification approach for feature selection in machine learning. It is a simple method for identifying the most relevant features in a dataset. See Table 1 for standardized coefficients of the strongest predictors for the logistic regression. The top predictors included:

1.3.2. Education

Dichotomous variable where 1 graduated high school or less, and 2 graduated some college.

1.3.3. Age

Participants were asked about their current age in years.

1.3.4. Friend support

Participants were asked four questions about support from their friends, such as "How much do your friends really care about you?" and "How much can you open up to them if you need to talk about your worries?". All items were answered on a 4-point Likert scale (1=*often*; 2=*sometimes*; 3=*rarely*; 4=*never*). Items were recoded so that higher scores reflected higher support. Internal consistency in the current sample was .88.

1.3.5. Waist-to-hip ratio

This variable was calculated by dividing the waist size (in inches) by the hip size (in inches). Higher ratios can mean more fat around the waist.

1.3.6. Depressed affect

Participants were asked seven questions about depressed affect during a two week period in the past 12 months. Items included "feel down on yourself, no good, or worthless?" and "lose interest in most things." The total score was constructed by adding the number of "Yes" responses.

1.3.7. Number of visits with medical doctors (in the past 12 months)

A continuous variable based on the total number of times the respondent reported seeing doctors for various reasons.

1.3.8. Number of sessions with a Mental Health Professionals (in the past 12 months)

A continuous variable based on the total number of times the respondent reported seeing professionals for emotional or mental health.

Table 1
The strongest predictors discriminating between recovery and nonrecovery as indicated by the logistic regression.

Predictor variables	Elastic net SC
Education	−3.96
Age	−1.57
Friend support	−1.57
Waist-to-hip-ratio	1.30
Depressed affect	1.26
No. of times seeing a medical health professional	1.00
No. of times seeing a mental health professional	0.96
Positive affect	−0.92
Family affectual solidarity	−0.91
Daily discrimination	0.78
Life satisfaction	−0.77

Note: Predictor variables are shown in order of importance (SC= standardized coefficients of log odds ratios).

1.3.9. Positive affect

Participants were asked six questions about how they felt during the past 30 days. Items included "cheerful?", "in good spirits?", and "full of life?". All items were answered on a five-point Likert scale (1=*all of the time; 2 =most of the time; 3 =some of the time; 4 =a little of the time;* 5 =*none of the time*). Items were recoded so that higher scores reflected higher levels of positive affect. Internal consistency in the current sample was .88.

1.3.10. Family affectual solidarity

This scale captured the degree of positive (and negative) sentiment between family members. It combined four items on family support (e.g., "How much can you rely on them for help if you have a serious problem") and four family strain items ("Not including your spouse or partner, how often do members of your family make too many demands on you"). Items for the family support subscale were recoded such that a high score signified high levels of family affectual solidarity. The total scale score was constructed by calculating the mean of the eight items measured on a 4-point Likert scale (1 = *often,* 4 = *never*). Internal consistency in the current sample was .65.

1.3.11. Life satisfaction

Respondents were asked to rate their life overall, work, health, relationship with spouse/partner, and relationship with children. Each item was rated from 0 (the worst possible) to 10 (the best possible). Items for relationship with spouse/partner and relationship with children were averaged to create a family relationship score. This score and the remaining three items were used to calculate an overall mean score. Higher scores reflected higher levels of overall life satisfaction. Internal consistency in the current sample was .66

1.3.12. Daily discrimination

Participants were asked nine questions about discrimination. Items included "You are treated with less courtesy than other people", "You are treated with less respect than other people", and "You receive poorer service than other people at restaurants or stores." All items were answered on a four-point Likert scale (1=*often and* 4=*never*). Items were recoded so that higher scores reflected higher levels of discrimination. Internal consistency in the current sample was .90.

1.3.13. Main data analysis

Two machine learning algorithms were compared to predict recovery or nonrecovery at follow-up. All analyses were conducted in Python using the scikit-learn library. Gradient boosted trees was selected over other machine learning algorithms because it incorporates regularization techniques to prevent overfitting (Bentéjac et al., 2021; Boehmke & Greenwell, 2019). This method assembles weak prediction models through continuous feature splitting and the addition of new trees to generate a more accurate model. All the trees are connected in sequence, with each tree attempting to reduce the error of the previous tree. The final model combines the results of each stage, producing a strong learner.

Penalized (elastic-net) logistic regression employs ridge regularization methods by decreasing model coefficients toward zero (Tikhonov, 1963), as well as least absolute shrinkage and selection operator (Lasso) regression (which reduces certain coefficients to zero; Tibshirani, 1996). Therefore, elastic net penalty enables regularization via ridge penalty with feature selection of the lasso penalty (Boehmke & Greenwell, 2019; James et al., 2013). It is a linear regression approach incorporating two penalty terms into the standard least-squares objective function. These two penalty terms represent the coefficient vector's L1 and L2 norms multiplied by two hyperparameters, alpha and lambda. The L1 norm is utilized for feature selection, and the L2 norm for feature shrinkage (Zou

& Hastie, 2005). Elastic net was selected because, in contrast to less transparent machine learning techniques (e.g., neural networks), it can preserve clinical interpretability and has a track record of reliability and accuracy (Ogutu et al., 2012; Zou & Hastie, 2005). Elastic net produces coefficients equivalent to log odds ratios of logistic regression. The models' hyperparameters were tuned with grid search, which involves training the models with different combinations of hyperparameters to find the best performance.

Comparing two machine learning algorithms, elastic net, and gradient boosted trees, is important in predicting recovery or non-recovery when considering the no-free lunch theorem. This theorem states that no universally superior algorithm in machine learning can solve all types of problems. Performance of an algorithm depends on the specific problem at hand and the data (Gómez & Rojas, 2016). Therefore, the theorem does not intend to find the perfect model to solve all problems but instead to find the best solution for a particular dataset and specific problem. Given the multifaceted nature of mental health prediction, it is important to explore multiple machine learning algorithms. By evaluating the elastic net, a linear model, and gradient boosted trees, a non-linear ensemble algorithm, we ensured a comprehensive examination of potential solutions in line with the theorem.

There is a risk of overfitting in machine learning studies with small sample sizes because the model can learn noise instead of patterns in the data. Nested cross-validation mitigates this risk by minimizing biased estimates of the true error (Krstajic et al., 2014; Varma & Simon, 2006). This method provides unbiased performance by iteratively separating the data used for hyperparameter tuning from data used for performance evaluation (Mueller & Guido, 2017). The approach also ensures a more accurate assessment of a model's generalizability (Varma & Simon, 2006). Nested cross-validation is beneficial when there is a limited quantity of data since it enables multiple training and testing of a model utilizing non-overlapping subsamples (i.e., folds) of the full dataset. Unlike traditional machine learning models that divide the sample into one training and one testing sample, this method uses the entire dataset for both training and testing. The dataset was split into five outer and three inner folds (Cawley & Talbot, 2010). The model is trained on each outer training fold and tested using the dataset withheld from that outer fold. The "inner loop" uses a grid search for parameter training and hyperparameter selection, with performance averaged over three folds. This iteration is repeated for all five outer folds, and the combined predictions from unseen test data across the inner folds were compared with the actual outcomes for the outer test folds. All performance measures provide an average across the respective folds. Once the model's performance was evaluated across all test partitions, a final step was performed by retraining the model with the optimal set of hyperparameters and selected predictors using the entire dataset.

Class imbalances can make machine learning outcomes skewed in favor of predicting the more prevalent class. We applied the borderline synthetic minority oversampling technique to handle these imbalances (SMOTE; Han et al., 2005). Borderline SMOTE generates new examples of the minority class using the nearest neighbors of these cases in the border region between cases (Han et al., 2005). To prevent data leakage, borderline SMOTE was used exclusively on the outer training folds, as balancing the entire dataset prior to cross-validation may lead to data leakage from the outer training folds to inner cross-validation hold-out test samples. In addition, data standardization was performed separately on the training and testing folds, so that the test set was not influenced by training data. Therefore, the test data was standardized without leakage from the training data.

Performance was measured with receiver operator characteristics (ROC) and area under the curve (AUC) metrics. An AUC value of .5 signifies no distinction between classes, whereas values greater than .5 indicate successful classification, effectively optimizing true positives and minimizing false positives. Although the interpretation of AUC values varies based on the specific classification task, we adhered to the generally accepted guidelines for AUC interpretation: AUC = .50 implies no distinction, AUC $\geq$ .70 and < .80 implies acceptable discrimination, and AUC $\geq$ .80 implies excellent discrimination (Mandrekar, 2010). Standard ROC metrics such as sensitivity (i.e., the proportion of positives correctly identified) and specificity (i.e., the proportion of correctly identified negatives) were evaluated. These metrics range between 0 and 1 such that larger values indicate better performance. Accuracy was estimated as the percentage of total items classified correctly.

2. Interpretability and variable importance

Interpretability in machine learning is needed so that humans can understand and analyze models for real-world applications (Molnar, 2019). Methods for explainable artificial intelligence were run using SHAP (Shapley Additive exPlanation) values to facilitate the interpretability of the models. These values draw inspiration from game theory principles and provide consistent and accurate measures of feature attribution (Marcílio & Eler, 2020). SHAP values assign a value to each variable that represents the average contribution of that variable across all possible combinations of variables. The average SHAP value across the sample is 0, but the average absolute SHAP value determines relative variable importance. The resulting SHAP values represent the importance or influence of each feature on the model's predictions. Since SHAP values are the industry standard for machine learning interpretability, they were chosen as the interpretability metric (Lundberg & Lee, 2017).

3. Results

Two models were tested: (i) gradient-boosted trees and (ii) elastic net. At nine-year follow-up, 95 participants (75.40 %) recovered, and 31 participants (24.60 %) did not recover.

Gradient boosted trees were trained on demographic, clinical, psychological, biological, and lifestyle predictors that discriminated between participants with and without a GAD diagnosis at 9-year follow-up with .62 AUC and 53 % balanced accuracy. As shown in Table 2, sensitivity was .34, and specificity was .85. The best model performance was associated with an Eta value of 0.1, depth of 3, and iterations of 200.

The elastic net trained on all predictors discriminated between participants with and without a GAD diagnosis at 9-year follow-up with .81 AUC and 72% balanced accuracy. Sensitivity was .70, and specificity was .76. The best model performance was associated with an alpha of 1.0 and a lambda value of 0.01. See Table 2 for performance metrics and tuning parameters.

3.1. Variable importance

Fig. 1 visually represents the top predictors of recovery or nonrecovery and displays their feature importance ranking for recovery (white) and nonrecovery (black) outcomes using mean absolute SHAP values. In order of feature importance, these included some college education or higher (recovery), older age (recovery), friend support (recovery), having higher waist-to-hip-ratio (recovery), daily discrimination (nonrecovery), higher positive affect (recovery), a greater number of sessions with a mental health professional in the past 12 months (nonrecovery), and a greater number of visits to medical doctors in the past 12 months (nonrecovery). The elastic net regularization shrunk family affectual solidarity and life satisfaction coefficients to zero, which suggests that these variables did not contribute significantly to the model's predictions.

4. Discussion

Our study compared two algorithms, gradient boosted trees and elastic net to predict recovery nine years later or not. Results suggested that prediction of individual recovery or nonrecovery of GAD was possible using supervised machine learning algorithms. The elastic net

Table 2
Performance metrics and tuning parameters for gradient boosted tree and elastic net models.

Algorithm	Sensitivity	Specificity	AUC	Balanced Accuracy	Tuning parameters		
Gradient Boosting	.34	.85	.62	.53	Eta= 0.1	Depth= 3	Iterations= 100
Elastic net	.70	.76	.81	.72	Alpha= 0.10	Lambda= 0.01	-

Note. AUC= Area under the curve; ROC= Receiver operating characteristic curve.

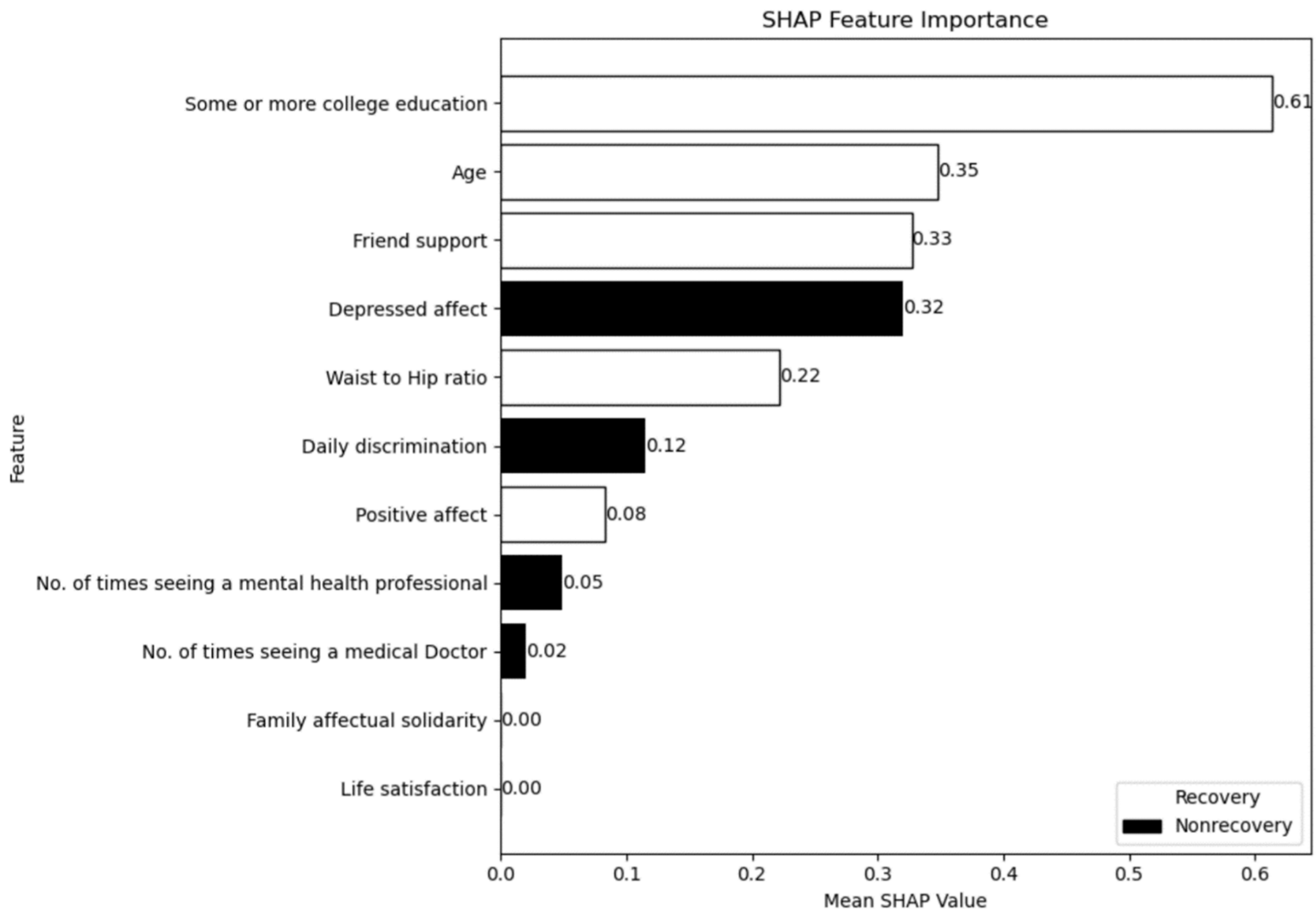

Fig. 1. Feature importance ranking for recovery (white bar) and nonrecovery (black bar) outcomes using mean absolute SHAP values.

model outperformed the gradient boosted trees model, achieving a balanced accuracy of 72 % with an AUC value of .81, which is considered excellent. To ensure robust and generalizable results, we employed nested cross-validation, which offers advantages over traditional validation methods like simple train-test split. By separating hyperparameter tuning from model evaluation, nested cross-validation prevents overfitting in machine learning studies with small sample sizes and reduces data leakage, resulting in an unbiased model performance (Mueller & Guido, 2017). This method is important for high-dimensional datasets with limited sample sizes, where standard methods can produce overly optimistic results (Lewis et al., 2023).

Overall, the elastic net algorithm revealed that some college education or more, older age, stronger support from friends, increased waist-to-hip ratio, and higher positive affect were associated with remission at nine-year follow-up. At the same time, depressed affect, daily discrimination, a greater number of sessions with a mental health professional, and a greater number of visits to medical doctors were associated with nonrecovery at a nine-year follow-up (see Fig. 1). Despite a wealth of research on factors predicting the onset of GAD, there is a notable gap in understanding predictors of remission, particularly within the context of GAD. Thus this study adds to those findings.

Having at least some college education was the most important protective factor associated with recovery according to SHAP feature importance. In particular, the mean absolute SHAP value was 0.61, indicating it had the greatest impact on the model's predictions (see Fig. 1). Higher levels of education have been linked to improved mental health in other studies (Belo et al., 2020; Niemeyer et al., 2019). Education is one of the most reliable predictors of life outcomes, including job, income, and socioeconomic status, and as such, may be viewed as an indicator of health and well-being (Galobardes et al., 2006; Javed & Khan, 2016). It may act as a protective factor by increasing receptivity to mental health messages and increasing the ability to communicate with mental healthcare providers to obtain appropriate psychotherapeutic treatments (Galobardes et al., 2006).

We also found that older age was the second strongest predictor of recovery at a nine-year follow-up, with a mean absolute SHAP value of 0.35. This is in line with previous studies suggesting anxiety symptoms tended to decrease over time or with advancing age (Kelly & Mezuk, 2017; Ramsawh et al., 2009). Possible mechanisms for a decrease in anxiety with advancing age include increased emotional control (Jorm, 2000). Research suggests that older adults report less experience of negative emotions (Gross et al., 1997) and a greater ability to control their emotions (Gross et al., 1997; Lawton et al., 1992). Thus, advancing age may be associated with increased ability to enhance positive emotions and dampen negative emotions (Lawton et al., 1992).

Friend support was pivotal at nine-year follow-up and (Gross et al., 1997) was the third strongest predictor of recovery, reflected by an absolute SHAP value of 0.33. It is well-established that social support plays an important role in safeguarding against anxiety (Lente et al., 2012; Zimmermann et al., 2020). For example, a study involving over 700,000 college students demonstrated that when faced with a disease outbreak, those who felt they had low levels of social support were 4.8 times more likely to experience anxiety and 6 times more likely to exhibit depressive symptoms, as compared to individuals who felt they had strong social support systems in place (Ma et al., 2020). Furthermore, positive social support was shown to act as a protective factor against the risk of emotional disorders by reducing stress and enhancing coping mechanisms (Cohen & Wills, 1985). Our study highlighted the important role of social support, encompassing personal relationships, in facilitating the recovery of GAD.

Greater depressed affect in those with GAD at baseline was the strongest predictor of nonrecovery at follow-up, demonstrated by a mean absolute SHAP of 0.32. Anxiety and depressive disorders are among the most common psychiatric illnesses, and they are highly comorbid with each other (Jacobson & Newman, 2017; Kessler et al., 2007). The lifetime comorbidity with depression was estimated to be around 59 % for individuals with GAD (Carter et al., 2001). Research suggests that patients with anxious major depressive disorder, compared with patients with nonanxious major depressive disorder, were found to have more severe role impairment (Baik et al., 2024; Kessler et al., 2015). Other data also suggest that those with comorbid anxiety and depression had a poorer course, greater functional impairment, greater disability, greater cardiovascular impairment, and more somatic morbidities (Ter Meulen et al., 2021). The influence of depressed affect on non-recovery underscores the importance of addressing depressive symptoms in the treatment of GAD.

Higher waist-to-hip ratio predicted recovery at a nine-year follow-up, with a mean absolute SHAP value of 0.22. This finding is at odds with prior research. A waist-to-hip ratio measure indicates abdominal body fat and has been linked to obesity (Gariepy et al., 2010; Milaneschi et al., 2019; Zhao et al., 2009). Depression and anxiety are two of the most common psychiatric disorders highly associated with obesity (Jorm et al., 2003; Strine et al., 2008). According to findings from a meta-analysis, the odds ratio (OR) for a connection between obesity and anxiety was 1.40 (confidence interval: 1.23–1.57; Gariepy et al., 2010). Rivenes et al. (2009) found that an elevated waist-to-hip ratio was concurrently linked to higher rates of anxiety and depression and other studies found a prospective relationship between obesity and later anxiety or GAD (Bjerkeset et al., 2008; Kasen et al., 2008). In our study, however, after accounting for body mass index (BMI), physical activity, social isolation, and somatic disorders, higher waist-to-hip ratio remained independently related to remission. Concurrently measured variables are marked by difficulty differentiating a particular variable's influence from confounding variables that may influence the study outcome. Our study controlled for other variables related to obesity such as BMI and lack of physical activity and it is therefore possible that without the influence of these factors, higher waist to hip ratio could be a more positive predictor of future GAD remission than expected.

Greater daily discrimination was the second strongest predictor associated with nonrecovery at follow-up, with a mean absolute SHAP of 0.12, underscoring its importance at nine-year follow-up. Research suggests that perceived discrimination has been linked to a wide range of adverse physical and mental health outcomes (Lewis et al., 2015; Williams et al., 2019). In particular, people who reported greater instances of discrimination had worse mental health (e.g., depression, psychiatric distress, and generalized anxiety disorder; Pieterse et al., 2012; Schmitt et al., 2014). People who face discrimination may be subjected to potentially damaging events, such as structural impediments to getting resources and interpersonal threats like ostracism or exclusion (Major et al., 2002). As a result, they may feel chronic stress due to the need to always be on alert against possible dangers (Lewis et al., 2015).

Positive affect was associated with recovery at follow-up, but it had less impact on recovery than other variables, with a mean absolute SHAP value of 0.08. Research suggests that positive affect is vital in maintaining mental health (Fredrickson, 2003; Rackoff & Newman, 2020). Positive affect may guard against the excessive avoidance that is the core of anxiety and depression (Fredrickson, 2003; Rackoff & Newman, 2020). For example, anxious individuals who reported positive emotions were less likely to engage in avoidance (Chow et al., 2017; Trew & Alden, 2012). Lastly, positive affect is believed to foster social connections, which lower the likelihood of mental disorders (Fredrickson, 2003; Jacobson & Lord, & Newman, 2017). This has led some to suggest that positive emotions set off "upward spirals" in mental health (Garland et al., 2010).

Greater number of sessions with a mental health professional in the past 12 months was a predictor of nonrecovery nine years later. However, it had a relatively small influence on the model's predictions of nonrecovery, as indicated by a mean absolute SHAP value of 0.05. Individuals with more consultations may have been struggling with more severe or treatment-resistant forms of GAD. One possible explanation is that these individuals may not have been receiving adequate evidence-based treatments, such as cognitive-behavioral therapy (CBT), which are effective in treating GAD (Borkovec et al., 2002; Newman et al., 2011). Alternatively, it could indicate that whereas evidence-based treatments were being provided, they may not have been fully effective for these individuals. Although CBT is considered a first-line treatment for GAD (Carpenter et al., 2018), it leads to substantial improvements in only about 50 % of people (Erickson & Newman, 2005; Newman et al., 2020).

The number of times seeing a medical professional in the past 12 months played a smaller but significant role in nonrecovery at follow-up, with a mean absolute SHAP value of 0.02. More medical visits likely indicated comorbid chronic medical conditions as those with general medical conditions in the past year were almost twice as likely to have a concurrent anxiety disorder (Sareen et al., 2005). In addition, those with both anxiety and general medical conditions tended to have significantly worse outcomes than those with either anxiety or general medical conditions alone (El-Gabalawy et al., 2011). For example, among older individuals with chronic conditions such as cataracts, allergies, and arthritis, those with anxiety reported worse mental and physical health (El-Gabalawy et al., 2011). Studies also revealed that individuals with both anxiety disorders and general medical conditions had a higher risk of suicidal ideation compared to those without a general medical condition (Raposo et al., 2014). In sum, the combination of general medical conditions and anxiety disorders leads to worse functioning compared to anxiety disorders or general medical conditions alone (Norman & Lang, 2005). Understanding these outcomes is essential, especially considering the prevalence of anxiety disorders in some general medical conditions (Norman & Lang, 2005).

These variables represent a range of psychological and interpersonal factors and offer valuable insights into potential recovery or nonrecovery for individuals with GAD. Understanding what predicts recovery from GAD allows for appropriate interventions and treatment planning. Although no model can definitively determine who will recover and who will not, these predictors serve as essential tools in assessing the likelihood of recovery from GAD.

The elastic net model had a false positive rate of 24 %, indicating that the model incorrectly classified about one-fourth of the actual negative cases as positive, and a false negative rate of about 30 %, indicating that they incorrectly classified about one-third of actual positive cases as negative. Several factors may account for this observation. MIDUS is a naturalistic cohort study where participants were subjected to varying environmental stressors and treatments throughout the nine-year follow-up period. These varying exposures may have influenced the outcomes over nine years.

To strengthen the generalizability of the findings, future research with larger sample sizes is required to ascertain whether the current results can be replicated. It should also aim to replicate similar outcomes using more diverse samples that encompass a wider variety of racial and ethnic groups. Additionally, the number of predictors in this study was relatively small compared to what is typically seen in machine learning studies that use larger datasets. Although our study utilized a comprehensive set of predictors from psychological, social, and biological domains, existing research has also demonstrated the effectiveness of physiological data in classifying anxiety disorders. Studies utilizing electroencephalography (EEG) and electrocardiography (ECG) have shown promising results in accurately classifying anxiety (Baygin et al., 2024; Sharma & Meena, 2024). These physiological measures provide valuable insights into anxiety's neural and cardiac correlates, offering an additional dimension that could enhance the predictive power of a machine learning model. Moreover, individual item-level data may be worth looking at as potential predictors to add to the traditional aggregate sum scores approach. Notably, prior machine learning research has demonstrated the potential of social media data, such as Twitter, to predict the onset and duration of emotional disorders, as well as neuroimaging data to predict treatment response (Reece et al., 2017; Shin et al., 2013). Predictive accuracy may improve by including these kinds of data and collecting data more frequently throughout the follow-up period. Research points to GAD following a chronic course, with retrospective studies suggesting that this chronic pattern can last up to 20 years (Bruce et al., 2005; Keller, 2002; Rickels & Schweizer, 1990). Although the current study focused on a 9-year follow-up, it may be necessary to consider the potential need for even longer tracking to fully understand the disorder's trajectory and impact on recovery and nonrecovery.

Several limitations should be noted. GAD was examined at baseline and then at a nine-year follow-up, which makes it difficult to determine the duration of GAD over the nine years. Moreover, the diagnostic interview focused on the previous six months and was not conducted annually over nine years. Therefore, even if individuals in the study had GAD at baseline and the nine-year follow-up, it is not possible to know how long they had been experiencing GAD symptoms across the 9 years. For example, it is possible that some individuals had GAD at baseline, remitted, and then experienced GAD symptoms again at the nine-year follow-up mark. As such, caution is advised when interpreting the findings as indicative of the chronic nature of GAD since the data only sheds light on recovery or nonrecovery at a nine-year follow-up, without offering insights into remission or recurrence of GAD between the two-time points when GAD was assessed

To the best of our knowledge, this is the first study using a machine learning algorithm to assess predictors of longterm recovery or nonrecovery in GAD. This represents an innovative approach in the field of anxiety disorders that has mostly relied on clinical judgment to predict risk for GAD recovery or nonrecovery. The study incorporated a wide range of predictors, most of which were previously related to the course of GAD at the group level. A strength of our study was the implementation of an interpretable machine-learning model that selected essential features to understand individual factors related to 9-year recovery or nonrecovery of GAD. The results of our study can be used as a benchmark for further research, which will probably improve and hone prediction powers. It has long been argued that statistical modeling will outperform clinician judgment in prediction tasks, and optimal predictive power is expected when statistical models and clinical interpretation are combined (Dwyer et al., 2018; Grove et al., 2000). As a result, clinical insights are increasingly being augmented by statistical models (Verma et al., 2021). As predictive models advance, they have the potential to aid in the development of secondary preventative measures and focused treatment options. Clinicians can utilize predictive models to personalize treatments based on expected outcomes and individual-specific factors. Predictive models can also be used to identify which individuals are likely to recover or are at risk of developing chronic symptoms. This enables the provision of relapse prevention resources, improving the overall efficiency of treatment. In the context of GAD, this study represents a noteworthy first step toward developing reliable machine learning-based predictions.

CRediT authorship contribution statement

Basterfield Candice: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Newman Michelle G.:** Writing – review & editing, Supervision, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of Competing Interest

The authors *declare* that they have no known competing *interests*.

Acknowledgments

Publicly available data from the MIDUS study was used for this research. Since 1995, the MIDUS study has been funded by the following: John D. and Catherine T. MacArthur: Foundation Research Network, National Institute on Aging (P01-AG020166), National Institute on Aging (U19-AG051426). This work was supported in part by National Institute of Mental Health Research Grant R01 MH115128.

References

American Psychiatric Association. (1980). *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed.).). American Psychiatric Association ⟨http://www.apa.org⟩.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; 5th ed.). American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596.

Baik, S. Y., Shin, K. E., Fitzsimmons-Craft, E. E., Eisenberg, D., Wilfley, D. E., Taylor, C. B., & Newman, M. G. (2024). The relationship of race, ethnicity, gender identity, sex assigned at birth, sexual orientation, parental education, financial hardship and comorbid mental disorders with quality of life in college students with anxiety, depression or eating disorders. *Journal of Affective Disorders, 366*, 335–344. https://doi.org/10.1016/j.jad.2024.08.098

Batelaan, N. M., Rhebergen, D., Spinhoven, P., van Balkom, A. J., & Penninx, B. W. (2014). Two-year course trajectories of anxiety disorders: Do DSM classifications matter? *Journal of Clinical Psychiatry, 75*(9), 3905–3909. https://doi.org/10.4088/JCP.13m08837

Baygin, M., Barua, P. D., Dogan, S., Tuncer, T., Hong, T. J., March, S., & Acharya, U. R. (2024). Automated anxiety detection using probabilistic binary pattern with ECG signals. *Computer Methods and Programs in Biomedicine, 247*, Article 108076. https://doi.org/10.1016/j.cmpb.2024.108076

Belo, P., Navarro-Pardo, E., Pocinho, R., Carrana, P., & Margarido, C. (2020). Relationship between mental health and the education level in elderly people: Mediation of leisure attitude. *Frontiers in Psychology, 11*, 1–7. https://doi.org/10.3389/fpsyg.2020.00573

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review, 54*(3), 1937–1967. https://doi.org/10.1007/S10462-020-09896-5

Bjerkeset, O., Romundstad, P., Evans, J., & Gunnell, D. (2008). Association of adult body mass index and height with anxiety, depression, and suicide in the general population: the HUNT study. *American Journal of Epidemiology, 167*(2), 193–202. https://doi.org/10.1093/aje/kwm280

Boehmke, B., & Greenwell, B. M. (2019). *Hands-on Machine Learning with R* (1st ed.).). Chapman and Hall/CRC. https://doi.org/10.1201/9780367816377

Bokma, W. A., Zhutovsky, P., Giltay, E. J., Schoevers, R. A., Penninx, B., van Balkom, A., & van Wingen, G. A. (2022). Predicting the naturalistic course in anxiety disorders using clinical and biological markers: A machine learning approach. *Psychological Medicine, 52*(1), 57–67. https://doi.org/10.1017/S0033291720001658

Borkovec, T. D., Newman, M. G., Pincus, A. L., & Lytle, R. (2002). A component analysis of cognitive-behavioral therapy for generalized anxiety disorder and the role of interpersonal problems. *Journal of Consulting and Clinical Psychology, 70*(2), 288–298. https://doi.org/10.1037/0022-006X.70.2.288

Bowes, S. M., Ammirati, R. J., Costello, T. H., Basterfield, C., & Lilienfeld, S. O. (2020). Cognitive biases, heuristics, and logical fallacies in clinical practice: A brief field guide for practicing clinicians and supervisors. *Professional Psychology: Research and Practice, 51*(5), 435–445. https://doi.org/10.1037/pro0000309

Bruce, S. E., Yonkers, K. A., Otto, M. W., Eisen, J. L., Weisberg, R. B., Pagano, M., & Keller, M. B. (2005). Influence of psychiatric comorbidity on recovery and recurrence in generalized anxiety disorder, social phobia, and panic disorder: A 12-

year prospective study. *American Journal of Psychiatry, 162*(6), 1179–1187. https://doi.org/10.1176/appi.ajp.162.6.1179

Calderon, A., Baik, S. Y., Ng, M. H. S., Fitzsimmons-Craft, E. E., Eisenberg, D., Wilfley, D. E., & Newman, M. G. (2024). Machine learning and Bayesian network analysis identifies psychiatric disorders and symptom associations with insomnia in a national sample of 31,285 treatment-seeking college students. *BMC Psychiatry, 24*(1), 656. https://doi.org/10.1186/s12888-024-06074-7

Carpenter, J. K., Andrews, L. A., Witcraft, S. M., Powers, M. B., Smits, J. A. J., & Hofmann, S. G. (2018). Cognitive behavioral therapy for anxiety and related disorders: A meta-analysis of randomized placebo-controlled trials. *Depression and Anxiety, 35*(6), 502–514. https://doi.org/10.1002/da.22728

Carter, R. M., Wittchen, H.-U., Pfister, H., & Kessler, R. C. (2001). One-year prevalence of subthreshold and threshold DSM-IV generalized anxiety disorder in a nationally representative sample. *Depression and Anxiety, 13*(2), 78–88. https://doi.org/10.1002/da.1020

Cawley, G. C., & Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research, 11*(70), 2079–2107. ⟨http://jmlr.csail.mit.edu/papers/v11/cawley10a.html⟩.

Chow, P. I., Fua, K., Huang, Y., Bonelli, W., Xiong, H., Barnes, L. E., & Teachman, B. A. (2017). Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. *Journal of Medical Internet Research, 19*(3), Article e62. https://doi.org/10.2196/jmir.6820

Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin, 98*(2), 310–357. https://doi.org/10.1037/0033-2909.98.2.310

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology, 14*, 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

El-Gabalawy, R., Mackenzie, C. S., Shooshtari, S., & Sareen, J. (2011). Comorbid physical health conditions and anxiety disorders: A population-based exploration of prevalence and health outcomes among older adults. *General Hospital Psychiatry, 33*(6), 556–564. https://doi.org/10.1016/j.genhosppsych.2011.07.005

Erickson, T. M., & Newman, M. G. (2005). Cognitive behavioral psychotherapy for generalized anxiety disorder: a primer. *Expert Review of Neurotherapeutics, 5*(2), 247–257. https://doi.org/10.1586/14737175.5.2.247

Fredrickson, B. L. (2003). The value of positive emotions: The emerging science of positive psychology is coming to understand why it's good to feel good. *American Scientist, 91*(4), 330–335. https://doi.org/10.1511/2003.4.330

Galobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W., & Smith, G. D. (2006). Indicators of socioeconomic position (part 1). *Journal of Epidemiology and Community Health, 60*(1), 7–12. https://doi.org/10.1136/jech.2004.023531

Gariepy, G., Nitka, D., & Schmitz, N. (2010). The association between obesity and anxiety disorders in the population: a systematic review and meta-analysis. *International Journal of Obesity, 34*(3), 407–419. https://doi.org/10.1038/ijo.2009.252

Garland, E. L., Fredrickson, B., Kring, A. M., Johnson, D. P., Meyer, P. S., & Penn, D. L. (2010). Upward spirals of positive emotions counter downward spirals of negativity: Insights from the broaden-and-build theory and affective neuroscience on the treatment of emotion dysfunctions and deficits in psychopathology. *Clinical Psychology Review, 30*(7), 849–864. https://doi.org/10.1016/j.cpr.2010.03.002

Gómez, D., & Rojas, A. (2016). An empirical overview of the no free lunch theorem and its effect on real-world machine learning classification. *Neural Computation, 28*(1), 216–228. https://doi.org/10.1162/NECO_a_00793

Gross, J. J., Carstensen, L. L., Pasupathi, M., Tsai, J., Skorpen, C. G., & Hsu, A. Y. (1997). Emotion and aging: Experience, expression, and control. *Psychology and Aging, 12*(4), 590–599. https://doi.org/10.1037//0882-7974.12.4.590

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*(1), 19–30. https://doi.org/10.1037/1040-3590.12.1.19

Gyorda, J. A., Nemesure, M. D., Price, G., & Jacobson, N. C. (2023). Applying ensemble machine learning models to predict individual response to a digitally delivered worry postponement intervention. *Journal of Affective Disorders, 320*, 201–210. https://doi.org/10.1016/j.jad.2022.09.112

Hahn, T., Nierenberg, A. A., & Whitfield-Gabrieli, S. (2017). Predictive analytics in mental health: Applications, guidelines, challenges and perspectives. *Molecular Psychiatry, 22*(1), 37–43. https://doi.org/10.1038/mp.2016.201

Han, H., Wang, W. Y., & Mao, B. H. (2005). Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In D.-S. Huang, X.-P. Zhang, & G.-B. Huang (Eds.), *Advances in intelligent computing, 3644* pp. 878–887). Springer. https://doi.org/10.1007/11538059_91.

Jacobson, N. C., Lord, K. A., & Newman, M. G. (2017). Perceived emotional social support in bereaved spouses mediates the relationship between anxiety and depression. *Journal of Affective Disorders, 211*, 83–91. https://doi.org/10.1016/j.jad.2017.01.011

Jacobson, N. C., & Newman, M. G. (2017). Anxiety and depression as bidirectional risk factors for one another: A meta-analysis of longitudinal studies. *Psychological Bulletin, 143*(11), 1155–1200. https://doi.org/10.1037/bul0000111

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). Statistical learning. In G. James, D. Witten, T. Hastie, & R. Tibshirani (Eds.), *An introduction to statistical learning: With applications in R, 103* pp. 15–57). Springer. https://doi.org/10.1007/978-1-4614-7138-7_2.

Javed, S., & Khan, A. (2016). Effect of education on quality of life and well being. *International Journal of Indian Psychology, 3*(3), 119–128.

Jenkins, D. G., & Quintana-Ascencio, P. F. (2020). A solution to minimum sample size for regressions. *PLoS ONE, 15*(2), Article e0229345. https://doi.org/10.1371/journal.pone.0229345

Jorm, A. F. (2000). Does old age reduce the risk of anxiety and depression? A review of epidemiological studies across the adult life span. *Psychological Medicine, 30*(1), 11–22. https://doi.org/10.1017/S0033291799001452

Jorm, A. F., Korten, A. E., Christensen, H., Jacomb, P. A., Rodgers, B., & Parslow, R. A. (2003). Association of obesity with anxiety, depression and emotional well-being: a community survey. *Australian and New Zealand Journal of Public Health, 27*(4), 434–440. https://doi.org/10.1111/j.1467-842X.2003.tb00423.x

Kasen, S., Cohen, P., Chen, H., & Must, A. (2008). Obesity and psychopathology in women: a three decade prospective study. *International Journal of Obesity, 32*(3), 558–566. https://doi.org/10.1038/sj.ijo.0803736

Keller, M. B. (2002). The long-term clinical course of generalized anxiety disorder. *Journal of Clinical Psychiatry, 63*(. 8), 11–16. ⟨www.psychiatrist.com⟩.

Kelly, K. M., & Mezuk, B. (2017). Predictors of remission from generalized anxiety disorder and major depressive disorder. *Journal of Affective Disorders, 208*, 467–474. https://doi.org/10.1016/j.jad.2016.10.042

Kessler, R. C., Merikangas, K. R., & Wang, P. S. (2007). Prevalence, comorbidity, and service utilization for mood disorders in the United States at the beginning of the twenty-first century. *Annual Review of Clinical Psychology, 3*, 137–158. https://doi.org/10.1146/annurev.clinpsy.3.022806.091444

Kessler, R. C., Sampson, N. A., Berglund, P., Gruber, M. J., Al-Hamzawi, A., Andrade, L., & De Girolamo, G. (2015). Anxious and non-anxious major depressive disorder in the World Health Organization World Mental Health Surveys. *Epidemiology and Psychiatric Sciences, 24*(3), 210–226. https://doi.org/10.1017/S2045796015000189

Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics, 6*(1), 10. https://doi.org/10.1186/1758-2946-6-10

Lawton, M. P., Kleban, M. H., Rajagopal, D., & Dean, J. (1992). Dimensions of affective experience in three age groups. *Psychology and Aging, 7*(2), 171–184. https://doi.org/10.1037//0882-7974.7.2.171

Lente, E., Barry, M., Molcho, M., Morgan, K., Watson, D., Harrington, J., & McGee, H. (2012). Measuring population mental health and social well-being. *International Journal of Public Health, 57*(2), 421–430. https://doi.org/10.1007/s00038-011-0317-x

Lewis, M. J., Spiliopoulou, A., Goldmann, K., Pitzalis, C., McKeigue, P., & Barnes, M. R. (2023). estedcv: an R package for fast implementation of nested cross-validation with embedded feature selection designed for transcriptomics and high-dimensional data. *Bioinformatics Advances, 3*(1), Article vbad048. https://doi.org/10.1093/bioadv/vbad048

Lewis, T. T., Cogburn, C. D., & Williams, D. R. (2015). Self-reported experiences of discrimination and health: scientific advances, ongoing controversies, and emerging issues. *Annual Review of Clinical Psychology, 11*(1), 407–440. https://doi.org/10.1146/annurev-clinpsy-032814-112728

Li, Y., Song, Y., Sui, J., Greiner, R., Li, X. M., Greenshaw, A. J., & Cao, B. (2024). Prospective prediction of anxiety onset in the Canadian longitudinal study on aging (CLSA): A machine learning study. *Journal of Affective Disorders, 357*, 148–155. https://doi.org/10.1016/j.jad.2024.04.098

Lueken, U., Zierhut, K. C., Hahn, T., Straube, B., Kircher, T., Reif, A., & Domschke, K. (2016). Neurobiological markers predicting treatment response in anxiety disorders: A systematic review and implications for clinical application. *Biobehavioral Reviews, 66*, 143–162. https://doi.org/10.1016/j.neubiorev.2016.04.005

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4768–4777. https://doi.org/10.5555/3295222.3295230

Ma, Z., Zhao, J., Li, Y., Chen, D., Wang, T., Zhang, Z., & Liu, X. (2020). Mental health problems and correlates among 746 217 college students during the coronavirus disease 2019 outbreak in China. *Epidemiology and Psychiatric Sciences, 29*, 1–10. https://doi.org/10.1017/S2045796020000931

Major, B., Gramzow, R. H., McCoy, S. K., Levin, S., Schmader, T., & Sidanius, J. (2002). Perceiving personal discrimination: the role of group status and legitimizing ideology. *Journal of Personality and Social Psychology, 82*(3), 269–282. https://doi.org/10.1037/0022-3514.82.3.269

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology, 5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Marcílio, W.E., & Eler, D.M. (2020). From explanations to feature selection: Assessing SHAP values as feature selection mechanism. Proceedings of the 2020 33rd SIBGRAPI conference on Graphics, Patterns and Images (SIBGRAPI), Porto de Galinhas, Brazil. 340-347. https://doi.org/10.1109/SIBGRAPI51738.2020.00053.

Martin, P. (2003). The epidemiology of anxiety disorders: A review. *Dialogues in Clinical Neuroscience, 5*(3), 281–298. https://doi.org/10.31887/DCNS.2003.5.3/pmartin

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. University of Minnesota Press. https://doi.org/10.1037/11281-000

Merikangas, K. R., & Pine, D. (2002). Genetic and other vulnerability factors for anxiety and stress disorders. In K. L. Davis, D. Charney, J. T. Coyle, & C. Nemeroff (Eds.), *Neuropsychopharmacology: The Fifth Generation of Progress* (pp. 867–882). Lippencott Williams & Wilkins.

Milaneschi, Y., Simmons, W. K., van Rossum, E. F., & Penninx, B. W. (2019). Depression and obesity: evidence of shared biological mechanisms. *Molecular Psychiatry, 24*(1), 18–33. https://doi.org/10.1038/s41380-018-0017-5

Mineka, S., & Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology of anxiety disorders: It's not what you thought it was. *American Psychologist, 61*(1), 10–26. https://doi.org/10.1037/0003-066X.61.1.10

Molnar, C. (2019). Interpretable machine learning. A Guide for Making\Black Box Models Explainable. Retrieved 7/27 from ⟨https://dlib.hust.edu.vn/bitstream/HUST/24999/1/OER000003227.pdf⟩.

Moons, K. G., Altman, D. G., Reitsma, J. B., Ioannidis, J. P., Macaskill, P., Steyerberg, E. W., & Collins, G. S. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine, 162*(1), W1–W73. https://doi.org/10.7326/M14-0698

Mueller, A., & Guido, S. (2017). *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media.

Newman, M. G., Basterfield, C., Erickson, T. M., Caulley, E., Przeworski, A., & Llera, S. J. (2022). Psychotherapeutic treatments for generalized anxiety disorder: Cognitive and behavioral therapies, enhancement strategies, and emerging efforts. *Expert Review of Neurotherapeutics, 22*(9), 751–770. https://doi.org/10.1080/14737175.2022.2125800

Newman, M. G., Castonguay, L. G., Borkovec, T. D., Fisher, A. J., Boswell, J., Szkodny, L. E., & Nordberg, S. S. (2011). A randomized controlled trial of cognitive-behavioral therapy for generalized anxiety disorder with integrated techniques from emotion-focused and interpersonal therapies. *Journal of Consulting and Clinical Psychology, 79*(2), 171–181. https://doi.org/10.1037/a0022489

Newman, M. G., Llera, S. J., Erickson, T. M., Przeworski, A., & Castonguay, L. G. (2013). Worry and generalized anxiety disorder: A review and theoretical synthesis of research on nature, etiology, and treatment. *Annual Review of Clinical Psychology, 9*(1), 275–297. https://doi.org/10.1146/annurev-clinpsy-050212-185544

Newman, M. G., Zainal, N. H., & Hoyer, J. (2020). Cognitive-behavioral therapy (CBT) for generalized anxiety disorder (GAD). In A. L. Gerlach, & A. T. Gloster (Eds.), *Generalized Anxiety Disorder and Worrying* (pp. 203–230). John Wiley & Sons. https://doi.org/10.1002/9781119189909.ch10.

Niemeyer, H., Bieda, A., Michalak, J., Schneider, S., & Margraf, J. (2019). Education and mental health: Do psychosocial resources matter? *SSM-Population Health, 7*, 1–12. https://doi.org/10.1016/j.ssmph.2019.100392

Norman, S. B., & Lang, A. J. (2005). The functional impact of anxiety sensitivity in the chronically physically ill. *Depression and Anxiety, 21*(4), 154–160. https://doi.org/10.1002/da.20076

Ogutu, J. O., Schulz-Streeck, T., & Piepho, H. P. (2012). Genomic selection using regularized linear regression models: Ridge regression, lasso, elasticnet and their extensions. *BMC Proceedings, 6*(. 2), Article S10. https://doi.org/10.1186/1753-6561-6-S2-S10

Penninx, B. W., Beekman, A. T., Smit, J. H., Zitman, F. G., Nolen, W. A., Spinhoven, P., & Van Der Meer, K. (2008). The Netherlands Study of Depression and Anxiety (NESDA): Rationale, objectives and methods. *International Journal of Methods in Psychiatric Research, 17*(3), 121–140. https://doi.org/10.1002/mpr.256

Penninx, B. W. J. H., Nolen, W. A., Lamers, F., Zitman, F. G., Smit, J. H., Spinhoven, P., & Beekman, A. T. F. (2011). Two-year course of depressive and anxiety disorders: Results from the Netherlands Study of Depression and Anxiety (NESDA). *Journal of Affective Disorders, 133*(1), 76–85. https://doi.org/10.1016/j.jad.2011.03.027

Pieterse, A. L., Todd, N. R., Neville, H. A., & Carter, R. T. (2012). Perceived racism and mental health among Black American adults: a meta-analytic review. *Journal of Counseling Psychology, 59*(1), 1–9. https://doi.org/10.1037/a0026208

Rackoff, G. N., & Newman, M. G. (2020). Reduced positive affect on days with stress exposure predicts depression, anxiety disorders, and low trait positive affect 7 years later. *Journal of Abnormal Psychology, 129*(8), 799–809. https://doi.org/10.1037/abn0000639

Radler, B. T., & Ryff, C. D. (2010). Who participates? Accounting for longitudinal retention in the MIDUS National Study of Health and Well-Being. *Journal of Aging and Health, 22*(3), 307–331. https://doi.org/10.1177/0898264309358617

Ramsawh, H. J., Raffa, S. D., Edelen, M. O., Rende, R., & Keller, M. B. (2009). Anxiety in middle adulthood: Effects of age and time on the 14-year course of panic disorder, social phobia and generalized anxiety disorder. *Psychological Medicine, 39*(4), 615–624. https://doi.org/10.1017/S0033291708003954

Raposo, S., El-Gabalawy, R., Erickson, J., Mackenzie, C. S., & Sareen, J. (2014). Associations between anxiety disorders, suicide ideation, and age in nationally representative samples of Canadian and American adults. *Journal of Anxiety Disorders, 28*(8), 823–829. https://doi.org/10.1016/j.janxdis.2014.09.005

Reece, A. G., Reagan, A. J., Lix, K. L., Dodds, P. S., Danforth, C. M., & Langer, E. J. (2017). Forecasting the onset and course of mental illness with Twitter data. *Scientific Reports, 7*(1), 1–10. https://doi.org/10.1038/s41598-017-12961-9

Rhebergen, D., Aderka, I. M., van der Steenstraten, I. M., van Balkom, A., van Oppen, P., Stek, M. L., & Batelaan, N. M. (2017). Admixture analysis of age of onset in generalized anxiety disorder. *Journal of Anxiety Disorders, 50*, 47–51. https://doi.org/10.1016/j.janxdis.2017.05.003

Rickels, K., & Schweizer, E. (1990). The clinical course and long-term management of generalized anxiety disorder. *Journal of Clinical Psychopharmacology, 10*(. 3), 101S–110S. https://doi.org/10.1097/00004714-199006001-00017

Rivenes, A. C., Harvey, S. B., & Mykletun, A. (2009). The relationship between abdominal fat, obesity, and common mental disorders: results from the HUNT study. *Journal of Psychosomatic Research, 66*(4), 269–275. https://doi.org/10.1016/j.jpsychores.2008.07.012

Rodriguez, B. F., Weisberg, R. B., Pagano, M. E., Bruce, S. E., Spencer, M. A., Culpepper, L., & Keller, M. B. (2006). Characteristics and predictors of full and partial recovery from generalized anxiety disorder in primary care patients. *Journal of Nervous and Mental Disease, 194*(2), 91–97. https://doi.org/10.1097/01.nmd.0000198140.02154.32

Sareen, J., Cox, B. J., Clara, I., & Asmundson, G. J. G. (2005). The relationship between anxiety disorders and physical disorders in the U.S. National Comorbidity Survey. *Depression and Anxiety, 21*(4), 193–202. https://doi.org/10.1002/da.20072

Schmitt, M. T., Branscombe, N. R., Postmes, T., & Garcia, A. (2014). The consequences of perceived discrimination for psychological well-being: a meta-analytic review. *Psychological Bulletin, 140*(4), 921–948. https://doi.org/10.1037/a0035754

Scholten, W. D., Batelaan, N. M., van Balkom, A. J., Penninx, B. W., Smit, J. H., & van Oppen, P. (2013). Recurrence of anxiety disorders and its predictors. *Journal of Affective Disorders, 147*(1-3), 180–185. https://doi.org/10.1016/j.jad.2012.10.031

Sharma, A., & Verbeke, W. J. (2021). Understanding importance of clinical biomarkers for diagnosis of anxiety disorders using machine learning models. *PLoS ONE, 16*(5), 1–7. https://doi.org/10.1371/journal.pone.0251365

Sharma, R., & Meena, H. K. (2024). Machine learning-based prediction of depression and anxiety using ECG signals. In R. J. Tripathy, & R. B. Pachori (Eds.), *Signal Processing Driven Machine Learning Techniques for Cardiovascular Data Processing* (pp. 65–80). Academic Press. https://doi.org/10.1016/b978-0-44-314141-6.00010-4.

Shin, L. M., Davis, F. C., VanElzakker, M. B., Dahlgren, M. K., & Dubois, S. J. (2013). Neuroimaging predictors of treatment response in anxiety disorders. *Biology of Mood Anxiety Disorders, 3*, 15. https://doi.org/10.1186/2045-5380-3-15

Spinhoven, P., Batelaan, N. M., Rhebergen, D., van Balkom, A. L., Schoevers, R., & Penninx, B. W. (2016). Prediction of 6-yr symptom course trajectories of anxiety disorders by diagnostic, clinical and psychological variables. *Journal of Anxiety Disorders, 44*, 92–101. https://doi.org/10.1016/j.janxdis.2016.10.011

Strine, T. W., Mokdad, A. H., Dube, S. R., Balluz, L. S., Gonzalez, O., Berry, J. T., & Kroenke, K. (2008). The association of depression and anxiety with obesity and unhealthy behaviors among community-dwelling US adults. *General Hospital Psychiatry, 30*(2), 127–137. https://doi.org/10.1016/j.genhosppsych.2007.12.008

Ter Meulen, W. G., Draisma, S., van Hemert, A. M., Schoevers, R. A., Kupka, R. W., Beekman, A. T. F., & Penninx, B. (2021). Depressive and anxiety disorders in concert-A synthesis of findings on comorbidity in the NESDA study. *Journal of Affective Disorders, 284*, 85–97. https://doi.org/10.1016/j.jad.2021.02.004

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological, 58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics, 4*, 1035–1038.

Trew, J. L., & Alden, L. E. (2012). Positive affect predicts avoidance goals in social interaction anxiety: Testing a hierarchical model of social goals. *Cognitive Behaviour Therapy, 41*(2), 174–183. https://doi.org/10.1080/16506073.2012.663402

Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. Article 91 *BMC Bioinformatics, 7*. https://doi.org/10.1186/1471-2105-7-91.

Verma, A. A., Murray, J., Greiner, R., Cohen, J. P., Shojania, K. G., Ghassemi, M., & Mamdani, M. (2021). Implementing machine learning in medicine. *Canadian Medical Association Journal, 193*(34), E1351–E1357. https://doi.org/10.1503/cmaj.202434

Williams, D. R., Lawrence, J. A., Davis, B. A., & Vu, C. (2019). Understanding how discrimination can affect health. *Health Services Research, 54*(2), 1374–1388. https://doi.org/10.1111/1475-6773.13222

Yonkers, K. A., Bruce, S. E., Dyck, I. R., & Keller, M. B. (2003). Chronicity, relapse, and illness-course of panic disorder, social phobia, and generalized anxiety disorder: Findings in men and women from 8 years of follow-up. *Depression and Anxiety, 17*(3), 173–179. https://doi.org/10.1002/da.10106

Yonkers, K. A., Dyck, I. R., Warshaw, M., & Keller, M. B. (2000). Factors predicting the clinical course of generalised anxiety disorder. *British Journal of Psychiatry, 176*(6), 544–549. https://doi.org/10.1192/bjp.176.6.544

Zainal, N. H., & Newman, M. G. (2024). Which client with generalized anxiety disorder benefits from a mindfulness ecological momentary intervention versus a self-monitoring app? Developing a multivariable machine learning predictive model. *Journal of Anxiety Disorders, 102*, Article 102825. https://doi.org/10.1016/j.janxdis.2024.102825

Zainal, N.H., & Newman, M.G. (in press). Prescriptive predictors of mindfulness ecological momentary intervention for social anxiety disorder: Machine learning analysis of randomized controlled trial data. *JMIR Mental Health*. https://doi.org/10.2196/67210.

Zhao, G., Ford, E.S., Dhingra, S., Li, C., Strine, T.W., Mokdad, A.H., & Depression and anxiety among US adults: associations with body mass index. International journal of obesity, 257-266. (2009). *Depression and anxiety among US adults: associations with body mass index. International Journal of Obesity*, *33*(2), 257-266. https://doi.org/10.1038/ijo.2008.268.

Zimmermann, M., Chong, A. K., Vechiu, C., & Papa, A. (2020). Modifiable risk and protective factors for anxiety disorders among adults: A systematic review. *Psychiatry Research, 285*, 1–16. https://doi.org/10.1016/j.psychres.2019.112705

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B, 67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x