Article

Assessment
2021, Vol. 28(8) 1971–1982

DOI: 10.1177/1073191120947797
journals.sagepub.com/home/asm
SAGE

Toward Reduced Burden in Evidence-Based Assessment of PTSD: A Machine Learning Study

Tammy Jiang[1], Sunny Dutra[1,2,3], Daniel J. Lee[1,2], Anthony J. Rosellini[1], Gabrielle M. Gauthier[2], Terence M. Keane[1,2], Jaimie L. Gradus[1], and Brian P. Marx[1,2]

Abstract

Structured diagnostic interviews involve significant respondent burden and clinician administration time. This study examined whether we can maintain diagnostic accuracy using fewer posttraumatic stress disorder (PTSD) assessment questions. Our study included 1,265 U.S. veterans of the Afghanistan and Iraq conflicts who were assessed for PTSD using the Structured Clinical Interview for the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (SCID-5). We used random forests to assess the importance of each diagnostic item in predicting a SCID-5 PTSD diagnosis. We used variable importance to rank each item and removed the lowest ranking items while maintaining ≥90% accuracy (i.e., efficiency), sensitivity, and other metrics. We eliminated six diagnostic items among the overall sample, four items among male veterans, and six items among female veterans. Our findings demonstrate that we may shorten the SCID-5 PTSD module while maintaining excellent diagnostic performance. These findings have implications for potentially reducing patient and provider burden of PTSD diagnostic assessment.

Keywords

posttraumatic stress disorder, machine learning, prediction, assessment, diagnosis

Identifying individuals with posttraumatic stress disorder (PTSD) is the first step to treatment and ultimately reducing the large societal and personal costs of this chronic and debilitating disorder (Kessler, 2000). Unfortunately, PTSD often goes undiagnosed. One reason PTSD may be missed in clinical settings is that providers often do not have adequate time to conduct a thorough mental health assessment (Ahmad et al., 2016; Beidas et al., 2015; Camara et al., 2000; Sheeran & Zimmerman, 2002). The development of brief and less arduous diagnostic procedures for PTSD would address such implementation barriers and increase the number of individuals who are properly identified with the disorder and who subsequently receive effective treatment (Beidas et al., 2015).

Structured diagnostic interviews are considered the optimal method for diagnosing PTSD because of their validity, reliability, and comprehensiveness (Weathers et al., 2009). However, these instruments require significant investments of institutional and clinician resources in training, administration, and interpretation and are often burdensome for patients with limited time and cognitive and emotional resources. For example, administration time for the PTSD module of the Structured Clinical Interview for the *Diagnostic Statistical Manual of Mental Disorders* (*DSM*), fifth edition (SCID-5) ranges from 30 to 40 minutes, and the Clinician-Administered Posttraumatic Stress Disorder Scale for *DSM-5* (CAPS-5) administration time ranges from 45 to 60 minutes (First et al., 2015; Spoont et al., 2013; Weathers et al., 2018) depending on the number of symptoms endorsed, openness and consistency of the respondent, and skill of the assessor. Increased burden limits the feasibility of administering these interviews in clinical settings where time for a thorough and comprehensive assessment is not available.

The long length of PTSD diagnostic assessment and the requirement of interview administration by a trained professional may contribute to delays in PTSD diagnosis or the

[1]Boston University, Boston, MA, USA
[2]VA Boston Healthcare System, Boston, MA, USA
[3]William James College, Newton, MA, USA

Corresponding Authors:
Jaimie L. Gradus, Department of Epidemiology, School of Public Health, Boston University, 715 Albany Street, T318E, Boston, MA 02118, USA.
Email: jgradus@bu.edu

Brian P. Marx, National Center for PTSD, VA Boston Healthcare System, 150 South Huntington Avenue, Boston, MA 02130, USA.
Email: brian.marx@va.gov

complete absence of PTSD assessments. Barriers to PTSD assessments translate into the lack of mental health care delivery, which is concerning because individuals with PTSD are at increased risk of a wide range of adverse health outcomes including physical disorders (Gradus et al., 2017, 2017; Jiang et al., 2018; Jiang et al., 2019) and suicide (Gradus et al., 2010). Thus, significant efforts have been invested in the design of abbreviated screening examinations with the goal of increasing rapid diagnoses for PTSD. Self-report PTSD measures, such as the Primary Care PTSD Screen for *DSM-5* (Prins et al., 2015) and PTSD Checklist for *DSM-5* (Blevins et al., 2015), are widely used in routine care to establish PTSD diagnostic status due to the limited response burden and ease of interpretation. However, these measures were never intended to serve this purpose. Consistent with the intended function of a screening measure, questionnaire measures often prioritize sensitivity to capture potential cases and have more limited ability to distinguish false positives. For example, the Primary Care PTSD Screen for *DSM-5* cutoff score of 3 had a sensitivity of 93% and specificity of 85% (Prins et al., 2016) relative to the MINI-International Neuropsychiatric Interview (Sheehan et al., 1998), and the PTSD Checklist for *DSM-5* cutoff score of 31 to 33 had a sensitivity of 81% and specificity of 71% relative to the CAPS-5 (Bovin et al., 2016). Accordingly, a measure that retains the diagnostic utility of a gold standard clinician-administered measure with less respondent burden would hold considerable value in routine clinical care. The aim of this article was to examine whether data-driven selection of questions in the PTSD module of the SCID-5 could result in an abbreviated and accurate instrument for the classification of PTSD.

Machine learning may enable us to identify subsets of items from a diagnostic interview that could facilitate faster diagnosis while maintaining validity using data-driven selection of diagnostic interview questions. Machine learning is a branch of computer science that aims to construct computer programs that automatically improve with experience (i.e., more data; Mitchell, 1997). Machine learning has been used to forecast PTSD after a traumatic event (Galatzer-Levy et al., 2014), identify sets of risk factors that increase the efficiency of early PTSD risk assessment (Karstoft et al., 2015), and develop biomarkers to predict subtypes of PTSD (Nicholson et al., 2019). There is a small, but growing body of research using machine learning to develop instruments that reduce the burden of using evidence-based assessments of mental disorders in clinical practice. Machine learning data reduction techniques are well suited for developing instruments that are both accurate and efficient by reducing the number of items required to accurately assess patients' symptoms. For example, one study used machine learning to develop an abbreviated version of a standardized diagnostic instrument for autism spectrum disorder, which reduced administration times (Wall et al., 2012). They found that eight of the 29 items contained in Module 1 of the Autism Diagnostic Observation Schedule—Generic were sufficient to classify autism with 100% accuracy. Another study abbreviated an assessment tool for pediatric obsessive-compulsive disorder and found that the machine learning algorithms reduced the number of items needed to identify obsessive-compulsive disorder diagnoses by 67% to 83% without sacrificing performance relative to the full subscales (Sattler et al., 2018). Stewart and colleagues (2016) developed a decision tree model using the chi-squared automatic interaction detector growth method to reduce the number of interview questions on the CAPS assessing *DSM-IV* symptom criteria for PTSD while maintaining diagnostic accuracy. They found that reducing the number of CAPS items administered by more than 75% could still retain 92% accuracy. Although these results are promising, it would not be clinically feasible to administer the items as suggested by their decision tree because they included functional impairment items as predictors. All structured diagnostic interviews of PTSD symptoms assess symptom-related distress and functional impairment only after the presence of symptoms has been assessed. Their model would require clinicians to ask respondents to describe functional impairment secondary to symptoms before the symptoms themselves are assessed, which is not possible. Furthermore, the method they used, decision trees, are vulnerable to overfitting, which occurs when a model is too specific to the data on which it is fit and is not useful for predicting events in new samples. An alternative approach would be to use a random forests classifier, which reduces the risk of overfitting because multiple trees are averaged together to produce more accurate and stable predictions compared with a single decision tree (Breiman, 2001a).

We used random forests in this study because of their ability to build prediction models that are accurate, robust, and interpretable (Hastie et al., 2009). Random forests have excellent performance compared with other classifiers (Breiman, 2001b; Brown & Mues, 2012; Maroco et al., 2011; Meyer et al., 2003; Statnikov et al., 2008; Svetnik et al., 2003) and are useful for assessing variable importance (Genuer et al., 2010), which may reveal items or constructs of the SCID-5 PTSD module that are most optimal at discriminating between persons with and without PTSD. This may inform the development of an abbreviated diagnostic tool focusing on the most important diagnostic items. We also conducted sex-stratified analyses because there are well-established sex differences in PTSD. Compared with men, women experience different types of trauma (e.g., sexual assault) more often, experience trauma at earlier ages, are more likely to develop PTSD after exposure to a traumatic event, report a greater number of PTSD symptoms, and report higher levels of certain PTSD symptoms such as exaggerated startle response and physiological cue activity (Birkeland et al., 2017; Breslau, 2002; Irish et al., 2011; Luxton et al., 2010). In this study, we explored the degree to which it is feasible to reduce the number of SCID-5 PTSD module items administered, while maintaining excellent diagnostic classification performance, using a machine learning approach both in an overall veteran sample and among men and women separately.

Method

Participants

Participants were enrolled in Project VALOR (Veterans' After-discharge Longitudinal Registry), a longitudinal study of veterans who were deployed in Operation Enduring Freedom and Operation Iraqi Freedom (Rosen et al., 2012). Project VALOR was designed to assess the natural history and outcomes associated with PTSD among veterans enrolled and receiving services in the Department of Veterans Affairs (VA) health care system. Veterans were eligible for Project VALOR if they were deployed in the conflicts in Afghanistan and Iraq, received a mental health evaluation at a VA facility between July 2008 and December 2009, and were not participating in a PTSD clinical trial at study initiation (Rosen et al., 2012). Veterans with probable PTSD (those who had at least two instances of a PTSD diagnosis by a mental health professional in separate visits) according to medical records were oversampled at a 3:1 ratio, and male and female veterans were sampled at a 1:1 ratio. Potential participants (n = 4,391) were contacted by phone and 2,712 consented to participate. Of the consented participants, 1,649 completed the first assessment. Doctoral-level clinicians conducted baseline interviews over the phone (Rosen et al., 2012), and we used these data to examine the demographic characteristics of the sample at baseline. Of the 1,649 participants who completed the baseline assessment, 1,347 participants completed the PTSD diagnostic interview at the second assessment approximately 2 years later. Persons who did not meet the trauma exposure criterion (n = 52) were not administered the SCID-5 PTSD module. In the current analyses, Project VALOR participants who were missing any of the SCID-5 items or had skipped too many of the items to determine whether they met the criteria for a PTSD diagnosis (n = 30) were excluded. This left 1,265 participants in our analytic sample. All study procedures were approved by the VA Boston Healthcare System Institutional Review Board and the Human Research Protection Office of the U.S. Army Medical Research and Materiel Command.

Measures

The SCID-5 is a semistructured interview for making the major *DSM-5* diagnoses (First et al., 2015). We administered the SCID-5 during the second assessment, which provided the data for our random forest models. We used the PTSD module of the SCID-5 to assess trauma history (Criterion A), one item each for the 20 core Criteria B to E symptoms, and one item each for symptom duration (Criterion F) and related distress or impairment in social, occupational, or other important areas of functioning (Criterion G,; First et al., 1998). Each symptom was rated as absent, subthreshold, or threshold based on interviewee responses to each item. Symptoms rated as "threshold" were considered present and counted toward the SCID-5 PTSD diagnosis. We did not use the skip out rules in the SCID-5 and assessed all PTSD symptoms even if the respondent did not fully meet criteria for PTSD diagnosis. This was done to obtain comprehensive symptom assessment for each participant. Interrater agreement was excellent (κ = .82) among a randomly selected subset of 100 interviews that were rated by a second assessor who separately listened to a digital recording of the interview and provided independent symptom and diagnosis ratings (Green et al., 2017).

Statistical Analyses

Random forests are an ensemble machine learning method that involves creating bootstrapped copies of the original data and training a classification tree on each bootstrapped copy, creating a collection of trees (Hastie et al., 2009). Each tree uses two thirds of the observations to train the model, and the remaining one third of the observations (i.e., out-of-bag observations) are used for testing the model (i.e., generating predicted values), which reduces overfitting since the model is tested on observations not used to train the model. Random forests also decorrelate the trees in the collection by reducing the number of predictors considered at each split of the tree. This is important because if there is one strong predictor, most or all of the trees in the collection will choose this variable to be the top split and all of the bagged trees will be similar to each other, creating highly correlated predictions (James et al., 2013). By forcing each split to consider only a subset of the predictors, weaker predictors will have more of a chance to be selected as split candidates.

We used the 20 symptoms included in PTSD Criteria B to E assessed by the SCID-5 PTSD module to predict the presence or absence of a SCID-5 PTSD diagnosis. Due to the aforementioned challenge of assessing related distress and impairment prior to symptoms, the items assessing clinically significant distress or impairment related to symptom severity were excluded from our statistical analyses as predictors. We built each random forest with 500 trees and determined the optimal number of predictors considered at each node by trying an increasing number of predictors considered at each split and then calculating the model's area under the receiver operating characteristic curve (AUC). We chose the number of predictors considered at each node that corresponded with the largest AUC value. We assessed predictor importance using Gini importance, a measure of node purity (Breiman, 2001a). A node is considered more "pure" if it contains predominantly observations from a single class (e.g., PTSD diagnosis or no PTSD diagnosis). The mean decrease in Gini measures how each variable contributes to the homogeneity of the nodes in the random forest. Variables that result in nodes with greater purity have larger decreases in mean Gini index and are thus more important for PTSD diagnostic classification.

Random forests were implemented using 10-fold cross validation to reduce risk of overfitting (Hastie et al., 2009). Diagnostic classification performance metrics, including accuracy (i.e., the proportion of the total number of predictions that were correct; also referred to as efficiency/correct classification rate), sensitivity (i.e., probability of a positive test result among PTSD cases), specificity (i.e., probability of a negative test result among subjects who do not have PTSD), positive predictive value (PPV; i.e., probability of being a PTSD case among subjects who have a positive test result), negative predictive value (NPV; i.e., probability of not having PTSD among subjects who have a negative test result), and AUC (i.e., ability of a test to distinguish between subjects with and without the disorder), were computed based on the predicted values estimated in the held-out folds. This procedure was repeated 10 times, and a different group of observations was treated as the validation set each time (Hastie et al., 2009; James et al., 2013). The model performance metrics (i.e., sensitivity, specificity, PPV, NPV, and AUC) from the 10 validation sets were then averaged to produce estimates of the model's predictive performance.

We constructed random forest models for the overall sample and stratified by sex. To determine whether it is possible to reduce the number of items on the SCID-5 PTSD module while preserving diagnostic classification performance, we removed items one at a time, starting from the least important item (item with the smallest mean decrease in Gini) and then removing the next least important item. After each removal of an item, we calculated the diagnostic classification performance measures. We used a cutoff of greater than .5 for the predicted probabilities to determine who was predicted to have PTSD. We removed as many items as we could while keeping model accuracy, sensitivity, specificity, PPV, NPV, and AUC values all above 0.90. Random forests were implemented via the caret package in the R statistical language (version 3.5.0; http://cran.r-project.org (Kuhn, 2008).

Results

Table 1 displays the demographic characteristics of the sample assessed at baseline. The mean age of the participants was 40.71 years (*SD* = 9.70). Roughly half of the participants were female, employed part or full time, married, or had an annual household income above $50,000. Most of the sample identified as White (78.24%) and heterosexual (92.06%). Approximately 43.07% of participants were college graduates. Table 2 displays the prevalence of comorbid psychiatric diagnoses in participants with and without PTSD diagnoses. Comorbid alcohol use disorder, depression, generalized anxiety disorder, and panic disorder were more prevalent among participants with PTSD than those without PTSD. Table 3 displays the prevalence of each PTSD symptom among the overall sample and among male and female participants separately. The prevalence of SCID-5 PTSD diagnoses was 65.70% in the overall sample, 65.30% among male participants, and 66.00% among female participants (Table 3). A major reason for why the prevalence of PTSD is so high in our sample is because we oversampled veterans who have PTSD, which was necessary for the aims of the parent project, Project VALOR.

Table 1. Characteristics of the Study Sample (n = 1,265).

Characteristic	*M* (*SD*) or percentage of total
Age, years, mean (*SD*)	40.71 (9.70)
Female (%)	50.99
Race (%)	
White	78.24
Black/African American	17.33
Hispanic/Latinx/Spanish	12.49
American Indian/Alaska Native	3.49
Asian	2.22
Native Hawaiian/Pacific Islander	0.71
College graduate (%)	43.07
Employed part or full time (%)	51.43
Annual household income ≥$50,000	49.19
Married (%)	52.78
Heterosexual (%)	92.06

Note. Participants were permitted to select multiple categories of ethnicity.

Table 2. Prevalence of Comorbid Diagnoses Among Participants With and Without PTSD.

Comorbid diagnosis	Participants with PTSD (%)	Participants without PTSD (%)
Alcohol use disorder	28.64	18.89
Depression	48.38	12.44
Generalized anxiety disorder	36.70	12.67
Panic disorder	44.77	14.06

Note. PTSD = posttraumatic stress disorder.

PTSD Prediction for Overall Sample

Figure 1 displays the variable importance plot for the overall sample. The least important item for distinguishing PTSD cases from noncases (diagnostic item with the lowest

Table 3. SCID-5 PTSD Symptom and PTSD Prevalence.[a]

PTSD symptom/diagnosis	Overall sample (n = 1,265)	Male participants (n = 620)	Female participants (n = 645)
B1: Intrusive memories	75.10	77.40	72.90
B2: Distressing dreams	56.40	56.50	56.40
B3: Dissociative reactions	19.90	19.20	20.60
B4: Cued psychological distress	61.00	61.10	60.90
B5: Cued physiological reactions	61.90	60.60	63.10
C1: Avoidance of memories, thoughts, feelings	70.70	69.20	72.10
C2: Avoidance of external reminders	78.20	75.80	80.50
D1: Inability to recall important aspect of event	23.90	19.80	27.80
D2: Exaggerated negative beliefs or expectations	63.60	61.60	65.60
D3: Distorted cognitions leading to blame	42.20	37.70	46.50
D4: Persistent negative emotional state	48.10	46.50	49.60
D5: Diminished interest or participation in activities	58.60	58.50	58.60
D6: Detachment or estrangement from others	70.90	70.20	71.60
D7: Persistent inability to experience positive emotions	39.40	41.30	37.70
E1: Irritable behavior and angry outbursts	58.80	57.60	60.00
E2: Reckless or self-destructive behavior	19.40	22.10	16.70
E3: Hypervigilance	86.00	87.40	84.70
E4: Exaggerated startle response	71.20	68.20	74.10
E5: Problems with concentration	68.80	68.90	68.70
E6: Sleep disturbance	72.60	70.20	74.90
PTSD diagnosis	65.70	65.30	66.00

Note. SCID-5 = Structured Clinical Interview for the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition; PTSD = posttraumatic stress disorder.
[a]All values are in percentage.

mean decrease in Gini) was dissociative reactions and the most important item was detachment or estrangement from others. Results indicated that questions assessing the following symptoms could be removed while maintaining all performance metrics greater than 0.90: dissociative reactions (B3), reckless or self-destructive behavior (E2), irritable behavior and angry outbursts (E1), hypervigilance (E3), persistent inability to experience positive emotions (D7), and exaggerated startle response (E4). After removing these items, the remaining items listed in the order of the greatest to the least variable importance for PTSD diagnostic classification were as follows: detachment or estrangement from others (D6); diminished interest or participation in activities (D5); avoidance of external reminders (C2); intrusive memories (B1); cued physiological reactions (B5); avoidance of memories, thoughts, and feelings (C1); exaggerated negative beliefs or expectations (D2); distressing dreams (B2); distorted cognitions leading to blame (D3); cued psychological distress (B4); persistent negative emotional state (D4); problems with concentration (E5); sleep disturbance (E6); and inability to recall important aspect of event (D1). The reduced model with these remaining 14 SCID-5 PTSD module items had an accuracy of 0.95, sensitivity of 0.97, specificity of 0.91, PPV of 0.95, NPV of 0.94, and AUC of 0.98 (see Table 4).

PTSD Prediction for Male Veterans

The variable importance measures for predicting PTSD among male veterans are shown in Figure 2. For male veterans, the items assessing the following symptoms could be removed while maintaining at least 0.90 on all diagnostic classification performance measures: inability to recall important aspect of event (D1), dissociative reactions (B3), reckless or self-destructive behavior (E2), and hypervigilance (E3). After removal of these items, the remaining items listed in order of the greatest to the least variable importance for PTSD diagnostic classification were as follows: detachment or estrangement from others (D6); diminished interest or participation in activities (D5); avoidance of external reminders (C2); exaggerated negative beliefs or expectations (D2); intrusive memories (B1); persistent negative emotional state (D4); problems with concentration (E5); cued physiological reactions (B5); cued psychological distress (B4); exaggerated startle response (E4); sleep disturbance (E6); avoidance of memories, thoughts, and feelings (C1); persistent inability to experience positive emotions (D7); distorted cognitions leading to blame (D3); distressing dreams (B2); and irritable behavior and angry outbursts (E1). The reduced model with the remaining 16 SCID-5 items had an accuracy of 0.93, sensitivity of 0.95, specificity of 0.91, PPV of 0.95, NPV of 0.90, and AUC of 0.97 (see Table 4).

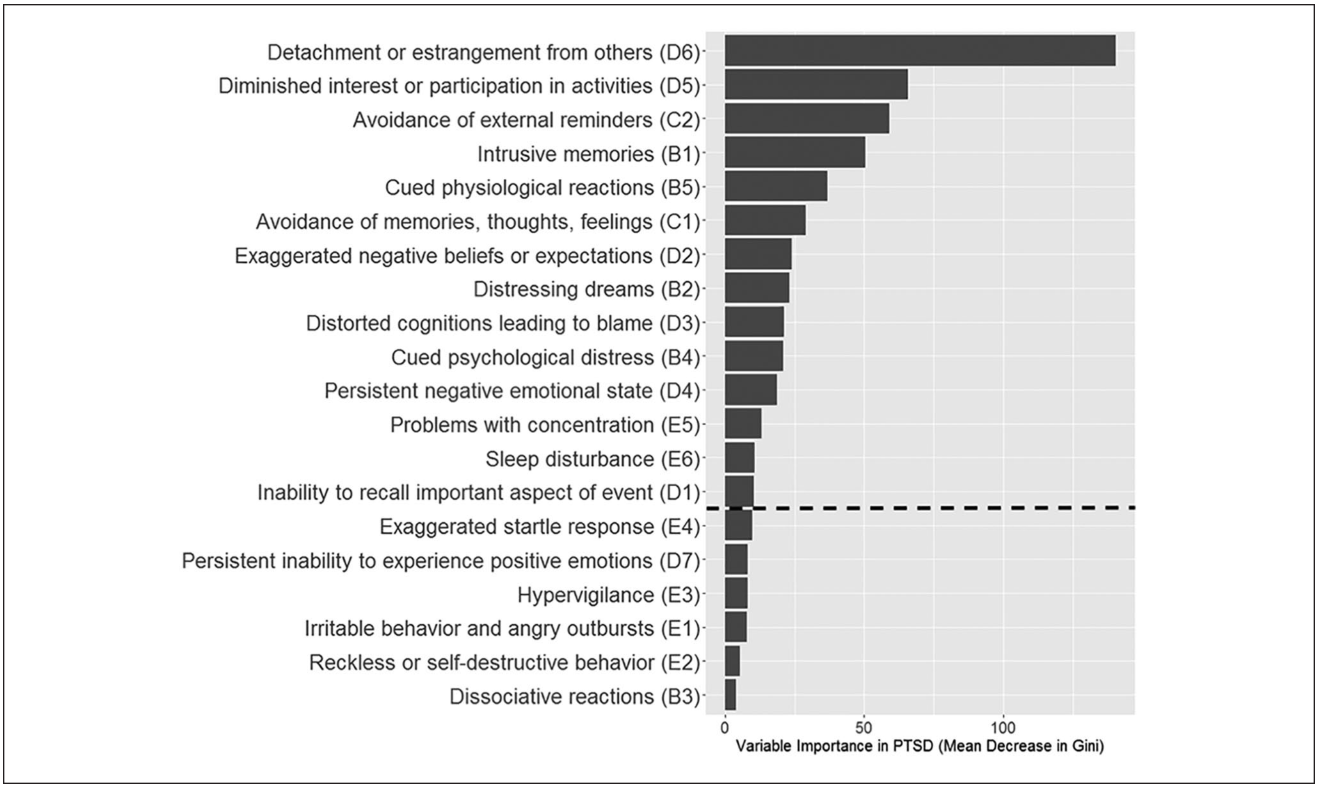

Figure 1. Variable importance for predicting PTSD among the overall sample.
Note. The *x*-axis shows the mean decrease in Gini, a measure of how each variable contributes to the purity of the nodes in the random forest. Lower values on the *x*-axis indicate lower variable importance for PTSD prediction. SCID items below the dashed line were excluded from the final reduced PTSD prediction model. PTSD = posttraumatic stress disorder; SCID = Structured Clinical Interview for the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition.

PTSD Prediction for Female Veterans

Figure 3 shows the variable importance plot for female veterans. For female veterans, the items assessing the following symptoms could be removed while maintaining at least 0.90 on all diagnostic classification performance measures: reckless or self-destructive behavior (E2), dissociative reactions (B3), persistent inability to experience positive emotions (D7), irritable behavior and angry outbursts (E1), exaggerated startle response (E4), and hypervigilance (E3). The remaining items listed in the order of the greatest to the least variable importance for PTSD diagnostic classification were as follows: detachment or estrangement from others (D6); intrusive memories (B1); avoidance of external reminders (C2); cued physiological reactions (B5); diminished interest or participation in activities (D5); avoidance of memories, thoughts, feelings (C1); exaggerated negative beliefs or expectations (D2); distressing dreams (B2); distorted cognitions leading to blame (D3); problems with concentration (E5); persistent negative emotional state (D4); cued psychological distress (B4); inability to recall important aspect of event (D1); and sleep disturbance (E6). The random forest constructed with the remaining 14 items had an accuracy of 0.95, sensitivity of 0.97, specificity of 0.90, PPV of 0.95, NPV of 0.93, and AUC of 0.99 (see Table 4).

The algorithm derived from the overall sample performs well in male and female participants, separately (Table 3). Among men, the accuracy of the overall sample algorithm among men was 94%, the probability of a false positive was 8.8%, and the probability of a false negative was 4%. Among women, the accuracy of the overall sample algorithm among women was 96%, the probability of a false positive was 9.6%, and the probability of a false negative was 1.9%.

Discussion

Current practices for diagnosing PTSD can be time consuming, expensive, and burdensome. The SCID-5 PTSD module, a diagnostic interview that is widely considered a gold standard diagnostic instrument in psychiatry, can take a long time to be administered and scored. This is a proof-of-concept study using machine learning to determine if we could develop an abbreviated, yet valid, diagnostic assessment protocol for PTSD that maintains validity. Our random forest algorithm in the overall sample of veterans reduced the number of items on

Table 4. Performance of Models After Removing the Most SCID-5 PTSD Module Items Possible While Keeping Accuracy, Sensitvity, Specificity, Positive Predictive Value, Negative Predictive Value, and AUC Greater Than 0.90.

Metric	Overall sample[a]	Male participants[b]	Female participants[c]
Accuracy (efficiency)	0.95	0.93	0.95
Sensitivity	0.97	0.95	0.97
Quality index of sensitivity	0.90	0.86	0.90
Specificity	0.91	0.91	0.90
Quality index of specificity	0.86	0.86	0.86
Positive predictive value	0.95	0.95	0.95
Negative predictive value	0.94	0.90	0.93
AUC	0.98	0.97	0.99

Note. The algorithm derived from the overall sample performs similarly in the sex-stratified samples. Among men, the accuracy of the overall sample algorithm among men was 94%, the probability of a false positive was 8.8%, and the probability of a false negative was 4%. Among women, the accuracy of the overall sample algorithm among women was 96%, the probability of a false positive was 9.6%, and the probability of a false negative was 1.9%. SCID-5 = Structured Clinical Interview for the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition; PTSD = posttraumatic stress disorder; AUC = area under the receiver operating characteristic curve.
[a]Excluded items B3, E2, E1, E3, D7, and E4. Included items D1, E6, E5, D4, B4, D3, B2, D2, C1, B5, B1, C2, D5, and D6. [b]Excluded items D1, B3, E2, and E3. Included items E1, B2, D3, D7, C1, E6, E4, B4, B5, E5, D4, B1, D2, C2, D5, and D6. [c]Excluded items E2, B3, D7, E1, E4, and E3. Included items E6, D1, B4, D4, E5, D3, B2, D2, C1, D5, B5, C2, B1, and D6.

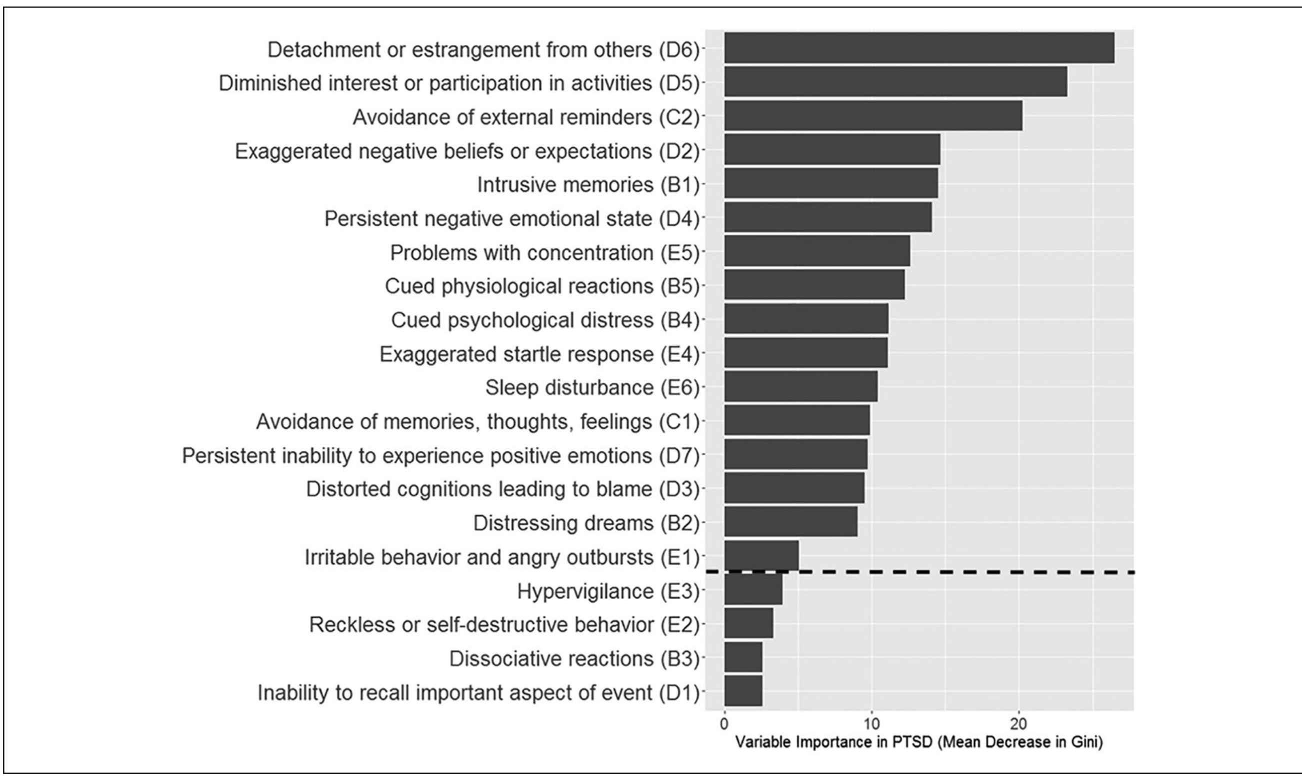

Figure 2. Variable importance for predicting PTSD among male participants only.
Note. The *x*-axis shows the mean decrease in Gini, a measure of how each variable contributes to the purity of the nodes in the random forest. Lower values on the *x*-axis indicate less variable importance for PTSD prediction. SCID items below the dashed line were excluded from the final reduced PTSD prediction model. PTSD = posttraumatic stress disorder; SCID = Structured Clinical Interview for the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition.

the SCID-5 PTSD module from 20 to 14 and had an accuracy of 0.95, sensitivity of 0.97, specificity of 0.91, PPV of 0.95, NPV of 0.94, and AUC of 0.98. Our findings may help guide future efforts to shorten the PTSD evaluation and diagnosis process such that individuals with PTSD may receive care sooner than under current diagnostic procedures.

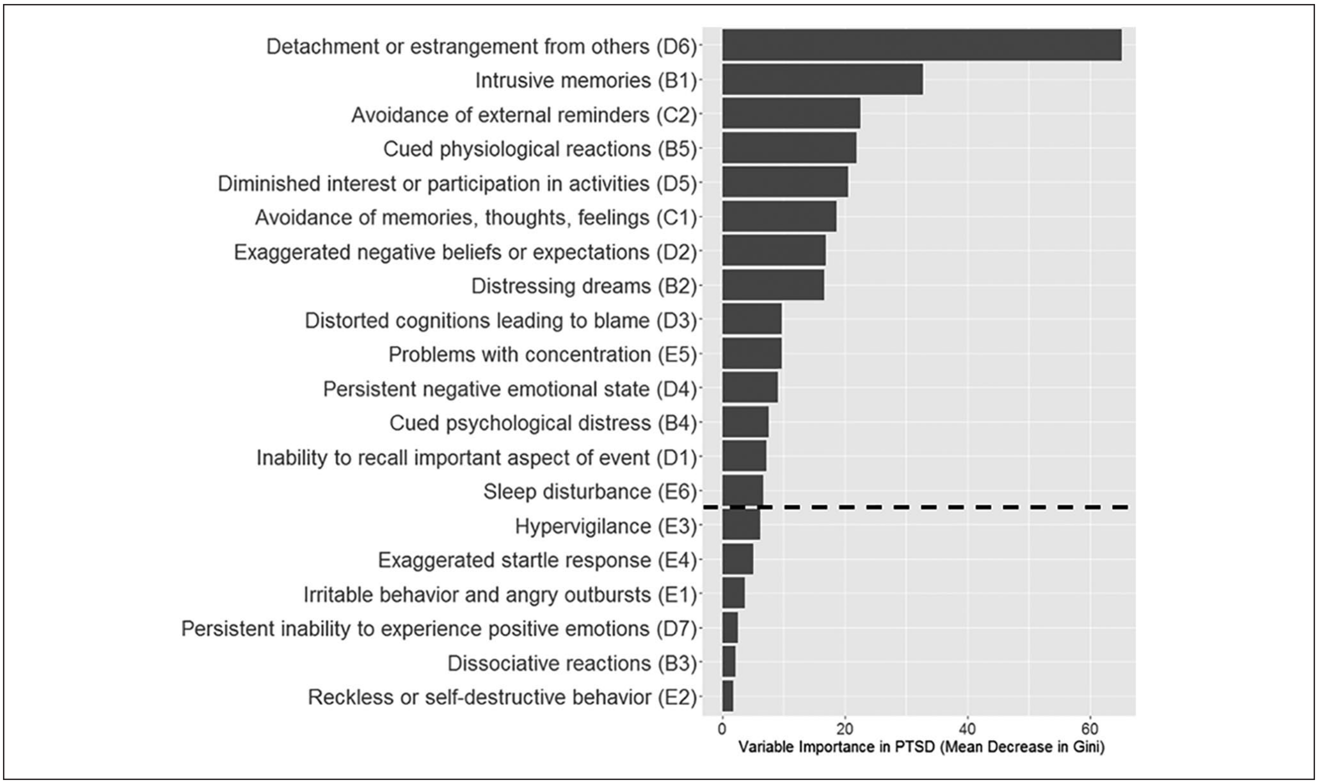

Figure 3. Variable importance for predicting PTSD among female participants only.
Note. The *x*-axis shows the mean decrease in Gini, a measure of how each variable contributes to the homogeneity of the nodes in the random forest. Lower values on the *x*-axis indicate lower variable importance for PTSD prediction. SCID items below the dashed line were excluded from the final reduced PTSD prediction model. The *x*-axis displaying the range of mean decrease in Gini values in Figures 1 to 3 differ because the mean decrease in Gini is affected by sample size such that the larger the sample is, the larger the mean decrease in Gini. The formula for calculating the mean decrease in Gini takes into account the size of a node (a point of the tree that splits into two), and larger samples will have larger node sizes. This explains why the maximum of the *x*-axis for the overall sample, including both men and women (Figure 1), is much larger than the maximums of the *x*-axes for men and women separately (Figures 2 and 3). PTSD = posttraumatic stress disorder; SCID = Structured Clinical Interview for the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition.

Clinician and respondent time burden are frequently cited as the most important factors prohibiting more widespread use of gold standard structured diagnostic interviews. Although we were able to reduce the number of SCID items required in order to make accurate diagnostic conclusions, the number of items removed was small. This advancement does not entirely solve the challenge imposed by additional time burden. However, assuming equal time across items, a 30% reduction in clinician and respondent time burden has potential value, particularly when considered in the context of routine care in a large health care system, such as the Veterans Health Administration (VHA). PTSD is one of the most common conditions for which veterans receive disability-related compensation in the VHA, with more than 1 million veterans receiving PTSD disability-related benefits from the VA (Veterans Benefits Administration Office of Performance Analysis and Integrity, 2019). Veterans must undergo a PTSD evaluation to receive this status, but the vast majority of evaluations do not involve standardized diagnostic assessments such as the SCID-5 because they are too time-consuming and burdensome (Jackson et al., 2011). A study that evaluated the extent to which veterans' PTSD disability benefit status corresponded to their PTSD diagnostic status, as determined by the SCID conducted by an independent assessor, found that the concordance between PTSD disability status and current and lifetime PTSD diagnosis on the SCID was 70% and 77%, respectively (Marx et al., 2016). These findings suggest that roughly 23% to 30% of veterans being considered for PTSD-related disability benefits are misclassified, indicating that a large proportion of health care resources may be misallocated. Some veterans with PTSD may not be receiving services that they need and others without PTSD may be receiving services they do not need.

Using a tool that automatically applies an accurate machine learning algorithm to a smaller set of symptom ratings may have the potential to reduce the burden of PTSD diagnostic assessment. If VHA clinicians who provide routine care were able to make all diagnostic conclusions 30% faster, this would free up clinicians to potentially see more patients over time. This is particularly important when considering the time burdens faced by health care institutions across the country. In clinical practice, computer-based diagnostic algorithms derived from machine learning methods could be disseminated and implemented to accelerate the pace of evaluation and ensure accurate PTSD assessment outcomes in settings that typically do not offer gold standard diagnostic interviews due to time and resource constraints (Bourla et al., 2018). Hand scoring of a random forest algorithm is not possible due to the complexity of the model (e.g., growing 500 trees, averaging across them, and performing 10-fold cross validation). Instead, clinicians would enter the 14 SCID patient symptom ratings into a desktop or app-based tool that could automatically apply the random forest model to the clinician ratings and instantaneously return the predicted probability of PTSD for each patient. Given that many clinicians are likely already using electronic systems, it may be feasible to integrate a tool that could be made widely accessible by clinicians assessing PTSD (e.g., in VA settings; McCarthy et al., 2015; Pan et al., 2017; Sattler et al., 2018; Wall et al., 2012; Wshah et al., 2019). Collectively, although this study was able to make a relatively small reduction in clinician and respondent time, the challenge of reducing this practical impediment to more valid assessment practices remains a critical hurdle for the field. Additional research is needed to develop an abbreviated diagnostic tool that reduces the administration time of the PTSD module, improves ease of administration and scoring, reduces clinician and patient time devoted to assessment, reduces waiting times for patients, and facilitates PTSD detection and linkage to treatment.

We used a large sample with equal proportions of men and women to examine how an abbreviated PTSD diagnostic algorithm may be different for men and women veterans. The sex-stratified random forest models revealed that the optimal items to remove while maintaining excellent diagnostic classification performance varied by sex, possibly reflecting sex-specific psychobiological responses to trauma, perceptions of threat and loss of control, and other factors related to the trauma (Olff et al., 2007). The least important item for predicting PTSD among men was inability to recall important aspect of event (D1), whereas this item was retained in the model for women to preserve accuracy. Irritable behavior and angry outbursts (E1) and exaggerated startle response (E4) were among the least important items for predicting PTSD and excluded from the reduced model for women, but for men, they were important for preserving excellent PTSD diagnostic accuracy. These results are consistent with prior research demonstrating that women tend to report higher emotional cue reactivity and peritraumatic dissociation (Olff et al., 2007), whereas men are more likely to report reckless or self-destructive behavior (Murphy et al., 2019). When we applied the algorithm developed in the overall sample (which was approximately 50/50 male/female) to the sex-stratified samples, the performance metrics in the sex-stratified samples were similar to those of the overall sample, indicating that the algorithm derived in the overall sample can also accurately predict PTSD in men and women separately. Additional research is needed to understand reasons for sex differences in PTSD, potential sex differences in symptom reporting, and whether diagnostic instruments tailored to men and women will improve PTSD identification. Future work should also investigate whether the optimal algorithm for PTSD prediction differs by other patient characteristics as well, such as age, race/ethnicity, trauma type, and sociocultural settings.

Our approach identifies the most efficient way to arrive at diagnostic classification using only the 20 Criteria B to E symptoms among those who endorsed a Criterion A traumatic event. The result that diagnostic classification with a reduced number of Criteria B to E symptoms was highly accurate absent any consideration of related distress or impairment is striking for two reasons. First, Stewart and colleagues (2016) found that related impairment items were the most important of any in guiding diagnostic conclusions. Second, the criterion that symptoms cause clinically significant distress or impairment is widely recognized as a common threshold for differentiating between normative and abnormal psychiatric symptoms. One possible explanation for this finding is that any participants experiencing Criteria B to E symptoms identified through this model were, by definition, also experiencing clinically significant distress and impairment since a symptom must cause clinically significant distress or impairment in their daily life to count toward a PTSD diagnosis. The symptoms identified in this study as having the greatest role in determining diagnostic classification (e.g., detachment or estrangement from others, diminished interest or participation in activities, avoidance of external reminders) imply reduced social functioning. Nonetheless, the result that a reduced set of Criteria B to E symptoms is highly accurate in determining PTSD diagnostic classification absent any consideration of symptom duration or related distress and impairment is a striking finding worthy of further research.

Our study is limited by loss to follow up. There were 302 participants from the first assessment who did not participate in the second assessment of Project VALOR, and therefore, they were removed from our analytic sample. Our random forest algorithms may have been different if we had information on these participants. A second study limitation is that our results may not be generalizable to persons who are not enrolled in VA health care services. Furthermore, we oversampled for female veterans and veterans with PTSD (Rosen et al., 2012). Thus, future research with different samples of trauma survivors (e.g., veterans who are not enrolled in VA health care, nonveteran trauma survivors, active duty service members) should examine the degree to which our results can be replicated and generalizable in these samples. Pending independent replication, our results may be applicable in PTSD specialty clinics for veterans. A third study limitation is that we did not have an external data set to validate our random forests models. However, we used out-of-bag observations to estimate the predictive performance of our random forests. In addition, we used 10-fold cross validation to estimate predictive performance in the test data not used in training the random forests. Future work should attempt to replicate our findings in an external sample. A fourth limitation is that our results do not provide information on the optimal order for administering items. Fifth, it is important to contextualize our findings relative to the value of full-length diagnostic assessment. If the goal of an assessment is not solely to determine diagnostic status, but to understand the breadth of trauma-related symptoms, identify dimensional symptom severity, or examine low base-rate symptoms (e.g., hypervigilance, dissociative reactions), then full-length diagnostic interviews are required. As with all assessment practice, the goals of the assessment should be weighed carefully in determining the appropriate measures. Comprehensive clinical assessment of the full range of PTSD symptoms may be more useful than an abbreviated diagnostic tool in some situations. Finally, our study does not propose an alternative to *DSM-5* PTSD criteria, which is known to have a high level of symptom profile heterogeneity (Galatzer-Levy & Bryant, 2013). In this study, we sought to examine if PTSD (as measured by the SCID-5) could be accurately diagnosed using a subset of SCID-5 items, rather than the full measure, as doing so may be cumbersome in many clinical situations (e.g., routine clinical care in general mental health, primary care settings, or specialty care clinics that are overwhelmed by patient demand and need briefer yet still accurate means of determining PTSD diagnostic status). Finally, our findings should be interpreted in light of potential criterion contamination. However, each of the individual PTSD symptoms had the same opportunity to predict the criterion (i.e., overall diagnostic status), but they did not do so equally.

An accurate, abbreviated diagnostic instrument may be more feasible to use than full-length structured interviews in routine clinical practice where there are time and resource constraints, especially in settings that do not provide PTSD assessments or in settings that use short PTSD screeners with less than excellent classification quality. Our results suggest that it is possible to develop an abbreviated, highly accurate PTSD diagnostic instrument using machine learning. We found that the random forests algorithms with a reduced set of items of the PTSD module of the SCID-5 were able to perform with excellent sensitivity, specificity, accuracy, PPV, NPV, and AUC. Our findings may help inform future work to computerize diagnostic assessments using machine learning and shorten the evaluation and diagnosis process such that individuals with PTSD may receive care more efficiently. Additional research is needed to evaluate the diagnostic accuracy of an abbreviated SCID-5 PTSD module that omits the items deemed unimportant for accurate PTSD diagnostic classification according to the SCID-5 in this study in an independent sample for external validation, determine whether an even shorter instrument could be developed using other analytic methods (e.g., item response theory; different machine learning techniques) while maintaining accuracy, and evaluate the effect of this abbreviated instrument on reducing patient and clinician burden. The development of such a tool may ultimately help reduce the number of undiagnosed and/or misdiagnosed PTSD cases in resource-limited settings and facilitate connection to treatment and services for individuals living with PTSD.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Department of Defense grants: W81XWH-08-2-0100/W81XWH-08-2-0102 and W81XWH-12-2-0117/W81XWH-12-2-0121. Drs. Dutra and Lee were supported by National Institute of Mental Health (NIMH) award #5T32MH019836-16. Dr. Gradus and Ms. Jiang were supported by NIMH grants 1R01MH110453-01A1 and 1R01MH109507.

ORCID iD

Tammy Jiang https://orcid.org/0000-0002-6887-5706

References

Ahmad, F., Shakya, Y., Ginsburg, L., Lou, W., Ng, P. T., Rashid, M., Ferrari, M., Ledwos, C., & McKenzie, K. (2016). Burden of common mental disorders in a community health centre sample. *Canadian Family Physician*, *62*(12), e758-e766.

Beidas, R. S., Stewart, R. E., Walsh, L., Lucas, S., Downey, M. M., Jackson, K., Fernandez, T., & Mandell, D. S. (2015). Free, brief, and validated: Standardized instruments for low-resource mental health settings. *Cognitive and Behavioral Practice*, *22*(1), 5-19. https://doi.org/10.1016/j.cbpra.2014.02.002

Birkeland, M. S., Blix, I., Solberg, Ø., & Heir, T. (2017). Gender differences in posttraumatic stress symptoms after a terrorist attack: A network approach. *Frontiers in Psychology*, *8*, 2091. https://doi.org/10.3389/fpsyg.2017.02091

Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The Posttraumatic Stress Disorder Checklist for *DSM-5* (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress*, *28*(6), 489-498. https://doi.org/10.1002/jts.22059

Bourla, A., Mouchabac, S., El Hage, W., & Ferreri, F. (2018). e-PTSD: An overview on how new technologies can improve prediction and assessment of posttraumatic stress disorder (PTSD). *European Journal of Psychotraumatology*, *9*(Suppl. 1). https://doi.org/10.1080/20008198.2018.1424448

Bovin, M. J., Marx, B. P., Weathers, F. W., Gallagher, M. W., Rodriguez, P., Schnurr, P. P., & Keane, T. M. (2016). Psychometric properties of the PTSD Checklist for *Diagnostic and Statistical Manual of Mental Disorders*—Fifth Edition (PCL-5) in veterans. *Psychological Assessment*, *28*(11), 1379-1391. https://doi.org/10.1037/pas0000254

Breiman, L. (2001a). Random forests. *Machine Learning*, *45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Breiman, L. (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199-231. https://doi.org/10.1214/ss/1009213726

Breslau, N. (2002). Gender differences in trauma and posttraumatic stress disorder. *Journal of Gender-Specific Medicine: JGSM*, *5*(1), 34-40.

Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, *39*(3), 3446-3453. https://doi.org/10.1016/j.eswa.2011.09.033

Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice*, *31*(2), 141-154. https://doi.org/10.1037/0735-7028.31.2.141

First, M., Spitzer, R., Gibbon, M., & Williams, J. (1998). *Structured Clinical Interview for DSM-IV Axis I Disorders*. Biometrics Research Department, New York State Psychiatric Institute.

First, M., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2015). *Structured Clinical Interview for DSM-5—Research Version* (SCID-5 for *DSM-5*, Research Version; SCID-5-RV). American Psychiatric Association.

Galatzer-Levy, I. R., & Bryant, R. A. (2013). 636,120 Ways to have posttraumatic stress disorder. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *8*(6), 651-662. https://doi.org/10.1177/1745691613504115

Galatzer-Levy, I. R., Karstoft, K.-I., Statnikov, A., & Shalev, A. Y. (2014). Quantitative forecasting of PTSD from early trauma responses: A machine learning application. *Journal of Psychiatric Research*, *59*, 68-76. https://doi.org/10.1016/j.jpsychires.2014.08.017

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, *31*(14), 2225-2236. https://doi.org/10.1016/j.patrec.2010.03.014

Gradus, J. L., Farkas, D. K., Svensson, E., Ehrenstein, V., Lash, T. L., & Toft, H. S. (2017). Posttraumatic stress disorder and gastrointestinal disorders in the Danish population. *Epidemiology (Cambridge, Mass.)*, *28*(3), 354-360. https://doi.org/10.1097/EDE.0000000000000622

Gradus, J. L., Qin, P., Lincoln, A. K., Miller, M., Lawler, E., Sørensen, H. T., & Lash, T. L. (2010). Posttraumatic stress disorder and completed suicide. *American Journal of Epidemiology*, *171*(6), 721-727. https://doi.org/10.1093/aje/kwp456

Green, J. D., Annunziata, A., Kleiman, S. E., Bovin, M. J., Harwell, A. M., Fox, A. M. L., Black, S. K., Schnurr, P. P., Holowka, D. W., Rosen, R. C., Keane, T. M., & Marx, B. P. (2017). Examining the diagnostic utility of the DSM-5 PTSD symptoms among male and female returning veterans. *Depression and Anxiety*, *34*(8), 752-760. https://doi.org/10.1002/da.22667

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer-Verlag. https://www.springer.com/us/book/9780387848570

Irish, L. A., Fischer, B., Fallon, W., Spoonster, E., Sledjeski, E. M., & Delahanty, D. L. (2011). Gender differences in PTSD symptoms: An exploration of peritraumatic mechanisms. *Journal of Anxiety Disorders*, *25*(2), 209-216. https://doi.org/10.1016/j.janxdis.2010.09.004

Jackson, J. C., Sinnott, P. L., Marx, B. P., Murdoch, M., Sayer, N. A., Alvarez, J. M., Greevy, R. A., Schnurr, P. P., Friedman, M. J., Shane, A. C., Owen, R. R., Keane, T. M., & Speroff, T. (2011). Variation in practices and attitudes of clinicians assessing PTSD-related disability among veterans. *Journal of Traumatic Stress*, *24*(5), 609-613. https://doi.org/10.1002/jts.20688

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer-Verlag. https://www.springer.com/us/book/9781461471370

Jiang, T., Farkas, D. K., Ahern, T. P., Lash, T. L., Sørensen, H. T., & Gradus, J. L. (2019). Posttraumatic stress disorder and incident infections: A nationwide cohort study. *Epidemiology (Cambridge, Mass.)*, *30*(6), 911-917. https://doi.org/10.1097/EDE.0000000000001071

Jiang, T., Veres, K., Körmendiné Farkas, D., Lash, T. L., Toft Sørensen, H., & Gradus, J. L. (2018). Post-traumatic stress disorder and incident fractures in the Danish population. *Osteoporosis International*, *29*(11), 2487-2493. https://doi.org/10.1007/s00198-018-4644-0

Karstoft, K.-I., Galatzer-Levy, I. R., Statnikov, A., Li, Z., & Shalev, A. Y., & For members of the Jerusalem Trauma Outreach and Prevention Study (J-TOPS) group. (2015). Bridging a translational gap: Using machine learning to improve the prediction of PTSD. *BMC Psychiatry*, *15*(1), 30. https://doi.org/10.1186/s12888-015-0399-8

Kessler, R. C. (2000). Posttraumatic stress disorder: The burden to the individual and to society. *Journal of Clinical Psychiatry*, *61*(Suppl. 5), 4-12.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(1), 1-26. https://doi.org/10.18637/jss.v028.i05

Luxton, D. D., Skopp, N. A., & Maguen, S. (2010). Gender differences in depression and PTSD symptoms following combat exposure. *Depression and Anxiety*, *27*(11), 1027-1033. https://doi.org/10.1002/da.20730

Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I., & de Mendonça, A. (2011). Data mining methods in the prediction of dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Research Notes*, *4*, 299. https://doi.org/10.1186/1756-0500-4-299

Marx, B. P., Bovin, M. J., Szafranski, D. D., Engel-Rebitzer, E., Gallagher, M. W., Holowka, D. W., Schnurr, P. P., Rosen, R. C., & Keane, T. M. (2016). Validity of posttraumatic stress disorder service connection status in Veterans Affairs electronic records of Iraq and Afghanistan Veterans. *Journal of Clinical Psychiatry*, *77*(4), 517-522. https://doi.org/10.4088/JCP.14m09666

McCarthy, J. F., Bossarte, R. M., Katz, I. R., Thompson, C., Kemp, J., Hannemann, C. M., Nielson, C., & Schoenbaum, M. (2015). Predictive modeling and concentration of the risk of suicide: Implications for preventive interventions in the US Department of Veterans Affairs. *American Journal of Public Health*, *105*(9), 1935-1942. https://doi.org/10.2105/AJPH.2015.302737

Meyer, D., Leisch, F., & Hornik, K. (2003). The support vector machine under test. *Neurocomputing*, *55*(1), 169-186. https://doi.org/10.1016/S0925-2312(03)00431-4

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.

Murphy, S., Elklit, A., Chen, Y. Y., Ghazali, S. R., & Shevlin, M. (2019). Sex differences in PTSD symptoms: A differential item functioning approach. *Psychological Trauma: Theory, Research, Practice and Policy*, *11*(3), 319-327. https://doi.org/10.1037/tra0000355

Nicholson, A. A., Densmore, M., McKinnon, M. C., Neufeld, R. W. J., Frewen, P. A., Théberge, J., Jetly, R., Richardson, J. D., & Lanius, R. A. (2019). Machine learning multivariate pattern analysis predicts classification of posttraumatic stress disorder and its dissociative subtype: A multimodal neuroimaging approach. *Psychological Medicine*, *49*(12), 2049-2059. https://doi.org/10.1017/S0033291718002866

Olff, M., Langeland, W., Draijer, N., & Gersons, B. P. R. (2007). Gender differences in posttraumatic stress disorder. *Psychological Bulletin*, *133*(2), 183-204. https://doi.org/10.1037/0033-2909.133.2.183

Pan, I., Nolan, L. B., Brown, R. R., Khan, R., van der Boor, P., Harris, D. G., & Ghani, R. (2017). Machine learning for social services: A study of prenatal case management in Illinois. *American Journal of Public Health*, *107*(6), 938-944. https://doi.org/10.2105/AJPH.2017.303711

Prins, A., Bovin, M. J., Kimerling, R., Kaloupek, D. G., Marx, B. P., Pless Kaiser, A., & Schnurr, P. P. (2015). *The Primary Care PTSD Screen for DSM-5 (PC-PTSD-5)* [Measurement instrument]. https://www.ptsd.va.gov

Prins, A., Bovin, M. J., Smolenski, D. J., Marx, B. P., Kimerling, R., Jenkins-Guarnieri, M. A., Kaloupek, D. G., Schnurr, P. P., Kaiser, A. P., Leyva, Y. E., & Tiet, Q. Q. (2016). The Primary Care PTSD Screen for DSM-5 (PC-PTSD-5): Development and evaluation within a veteran primary care sample. *Journal of General Internal Medicine*, *31*(10), 1206-1211. https://doi.org/10.1007/s11606-016-3703-5

Rosen, R. C., Marx, B. P., Maserejian, N. N., Holowka, D. W., Gates, M. A., Sleeper, L. A., Vasterling, J. J., Kang, H. K., & Keane, T. M. (2012). Project VALOR: Design and methods of a longitudinal registry of post-traumatic stress disorder (PTSD) in combat-exposed veterans in the Afghanistan and Iraqi military theaters of operations. *International Journal of Methods in Psychiatric Research*, *21*(1), 5-16. https://doi.org/10.1002/mpr.355

Sattler, A. F., Whiteside, S. P. H., Bentley, J. P., & Young, J. (2018). Development and validation of a brief screening procedure for pediatric obsessive-compulsive disorder derived from the Spence Children's Anxiety Scale. *Journal of Obsessive-Compulsive and Related Disorders*, *16*, 29-35. https://doi.org/10.1016/j.jocrd.2017.12.004

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. *Journal of Clinical Psychiatry*, *59* (Suppl. 20), 22-33; quiz 34-57.

Sheeran, T., & Zimmerman, M. (2002). Screening for posttraumatic stress disorder in a general psychiatric outpatient setting. *Journal of Consulting and Clinical Psychology*, *70*(4), 961-966.

Spoont, M., Arbisi, P., Fu, S., Greer, N., Kehle-Forbes, S., Meis, L., Rutks, I., & Wilt, T. J. (2013). *Screening for post-traumatic stress disorder (PTSD) in primary care: A systematic review*. Department of Veterans Affairs. http://www.ncbi.nlm.nih.gov/books/NBK126691/

Statnikov, A., Wang, L., & Aliferis, C. F. (2008). A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinformatics*, *9*, 319. https://doi.org/10.1186/1471-2105-9-319

Stewart, R. W., Tuerk, P. W., Metzger, I. W., Davidson, T. M., & Young, J. (2016). A decision-tree approach to the assessment of posttraumatic stress disorder: Engineering empirically rigorous and ecologically valid assessment measures. *Psychological Services*, *13*(1), 1-9. https://doi.org/10.1037/ser0000069

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences*, *43*(6), 1947-1958. https://doi.org/10.1021/ci034160g

Veterans Benefits Administration Office of Performance Analysis and Integrity. (2019, November 5). *VA benefits & health care utilization*. National Center for Veterans Analysis and Statistics. https://www.va.gov/vetdata/docs/pocketcards/fy2020q1.PDF

Wall, D. P., Kosmicki, J., DeLuca, T. F., Harstad, E., & Fusaro, V. A. (2012). Use of machine learning to shorten observation-based screening and diagnosis of autism. *Translational Psychiatry*, *2*(4), e100. https://doi.org/10.1038/tp.2012.10

Weathers, F. W., Bovin, M. J., Lee, D. J., Sloan, D. M., Schnurr, P. P., Kaloupek, D. G., Keane, T. M., & Marx, B. P. (2018). The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5): Development and initial psychometric evaluation in military veterans. *Psychological Assessment*, *30*(3), 383-395. https://doi.org/10.1037/pas0000486

Weathers, F. W., Keane, T. M., & Foa, E. B. (2009). Assessment and diagnosis of adults. In *Effective treatments for PTSD: Practice guidelines from the International Society for Traumatic Stress Studies* (2nd ed., pp. 23-61). Guilford Press.

Wshah, S., Skalka, C., & Price, M. (2019). Predicting posttraumatic stress disorder risk: A machine learning approach. *JMIR Mental Health*, *6*(7), e13946. https://doi.org/10.2196/13946