



Published in final edited form as:

J Psychiatr Res. 2014 December ; 59: 68–76. doi:10.1016/j.jpsychires.2014.08.017.

Quantitative Forecasting of PTSD from Early Trauma Responses: A Machine Learning Application

Isaac R. Galatzer-Levy, Ph.D.², Karen-Inge Karstoft, Ph.D.^{2,3}, Alexander Statnikov, Ph.D.^{4,5}, and Arieh Y. Shalev, M.D.^{1,2}

¹Center for Traumatic Stress Studies, Hadassah University Hospital, Jerusalem, Israel

²Department of Psychiatry, NYU School of Medicine, New York, NY

³Department of Psychology, University of Southern Denmark, Odense, Denmark

⁴Center for Health Informatics and Bioinformatics, NYU School of Medicine, New York, NY

⁵Department of Medicine, NYU School of Medicine, New York, NY

Abstract

There is broad interest in predicting the clinical course of mental disorders from early, multimodal clinical and biological information. Current computational models, however, constitute a significant barrier to realizing this goal. The early identification of trauma survivors at risk of post-traumatic stress disorder (PTSD) is plausible given the disorder's salient onset and the abundance of putative biological and clinical risk indicators. This work evaluates the ability of Machine Learning (ML) forecasting approaches to identify and integrate a panel of unique predictive characteristics and determine their accuracy in forecasting non-remitting PTSD from information collected within 10 days of a traumatic event. Data on event characteristics, emergency department observations, and early symptoms were collected in 957 trauma survivors, followed for fifteen months. An ML feature selection algorithm identified a set of predictors that rendered all others redundant. Support Vector Machines (SVMs) as well as other ML classification algorithms were used to evaluate the forecasting accuracy of i) ML selected features,

© 2014 Elsevier Ltd. All rights reserved.

*Corresponding Author: Isaac R. Galatzer-Levy, Ph.D., New York University School of Medicine, 1 Park Ave, New York, NY 10016, 847-420-2527, Isaac.Galatzer-Levy@nyumc.org.

Member of the J-TOPS group: Yael Ankri, M.A.; Sara Freedman, Ph.D.; Rhonda Addeky, Ph.D.; Yossi Israeli – Shalev, M.A.; Moran Gilad, M.A.; Pablo Roitman M.D.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Financial Disclosures

AYS received an investigator-initiated grant from Lundbeck Pharmaceuticals for this work and for a collaborative study entitled "Prevention of PTSD by Escitalopram;" (principal investigator: Dr. Joseph Zohar).

The funders had no role in study design, data collection and analysis, manuscript preparation and publication.

Other authors have no financial disclosures to report.

Contributors

Isaac R. Galatzer-Levy: Isaac Galatzer-Levy was involved in writing and data analysis and interpretation of results.

Karen-Inge Karstoft: Karen-Inge Karstoft was involved in writing and data analysis and interpretation of results.

Alexander Statnikov: Alexander Statnikov was involved in data analysis and interpretation of results

Arieh Shalev: Arieh Shalev was involved in writing and data analysis and interpretation of results.

ii) all available features without selection, and iii) Acute Stress Disorder (ASD) symptoms alone. SVM also compared the prediction of a) PTSD diagnostic status at 15 months to b) posterior probability of membership in an empirically derived non-remitting PTSD symptom trajectory. Results are expressed as mean Area Under Receiver Operating Characteristics Curve (AUC). The feature selection algorithm identified 16 predictors, present in 95% cross-validation trials. The accuracy of predicting non-remitting PTSD from that set (AUC=.77) did not differ from predicting from all available information (AUC=.78). Predicting from ASD symptoms was not better than chance (AUC =.60). The prediction of PTSD status was less accurate than that of membership in a non-remitting trajectory (AUC=.71). ML methods may fill a critical gap in forecasting PTSD. The ability to identify and integrate unique risk indicators makes this a promising approach for developing algorithms that infer probabilistic risk of chronic posttraumatic stress psychopathology based on complex sources of biological, psychological, and social information.

Keywords

Posttraumatic Stress Disorder (PTSD); Course and Prognosis; Early Prediction; forecasting; Machine Learning; Markov boundary feature selection; Support Vector Machines

1. Introduction

Chronic PTSD is prevalent, distressful, and debilitating (Kessler, 2000) and often follows an unremitting course (Galatzer-Levy et al., 2013; Peleg & Shalev, 2006). The early manifestations may provide sufficient information to identify individuals at risk for chronic PTSD. Studies to date have identified numerous risk indicators of chronic PTSD, many of which are accessible shortly after trauma exposure. These include, but are not limited to early symptoms of PTSD, depression or dissociation, physiological arousal (e.g., heart rate), early neuroendocrine responses, gender, lower socio-economic status, the early use of opiate analgesics, the occurrence of traumatic brain injury, and a progressively growing number of genetic and transcriptional factors (Boscarino, Erlich, Hoffman, & Zhang, 2012; Brewin, Andrews, & Valentine, 2000; Etkin & Wager, 2007; Karl et al., 2006; Ozer, Best, Lipsey, & Weiss, 2003). Despite these discoveries, the individual identification of risk for PTSD remains elusive, thereby leaving a major gap between scientific discovery and practical application.

One reason for such a gap is the current use of computational models that do not match the disorder's inherent complexity in etiology. As attested by its numerous risk factors, the etiology of PTSD is multi-causal, multi-modal, and complex. As such, the longitudinal course of PTSD reflects a converging interaction of numerous, multimodal risk factors. Moreover, specific risk markers and their relative weight can vary between individuals and traumatic circumstances. For example, head injury increases the likelihood of developing PTSD (Bryan & Clemans, 2013) but has a low occurrence overall among survivors. Similarly, the contribution of female gender to the risk of developing PTSD varies between traumatic events (Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995) and with specific genetic risk alleles (Ressler et al., 2011). Thus, to accurately predict PTSD in individuals, one must account for complex and variable interactions between putative markers.

The commonly utilized General Linear Modeling (GLM) was designed to test focused hypotheses without generalization beyond the data under study (Hald, 2007). This modeling approach identifies the probability of rejecting a null hypothesis of no effect along with an estimate of shared variance between the dependent and independent variable(s) (Cohen, 1994). Such models are important for rigorously testing novel hypotheses because they assume no relationship between variables. In this context, a significant p-value indicates disconfirmation of the assumption of no relationship between variables. Further, estimates of unique and shared variance provide information about how much of the variability in the dependent variable (DV) is accounted for by the independent variable(s) (IVs). This approach has intrinsic limitations. First, it is built on the assumption that variables would follow a normal distribution given an infinite sample size (Stigler, 1986). Second, information provided is limited to null hypothesis testing along with effect size estimates of shared variance. Third, relatively large sample-to-variable ratios are needed for such analyses because relationships between variance components are being analyzed. What is needed to forecast later outcomes is the probability of the DV given the IV(s) along with an estimate of accuracy. Further, as many identified predictors may be redundant, methods to identify variables that provide unique information are also important. Finally, as many different variables may provide predictive information, analysis and integration of many variables simultaneously is required.

Machine learning (ML) can handle large complex data with heterogeneous distributions (Hastie, Tibshirani, & Friedman, 2003), determine probabilistic relationships from complex conditional dependencies between variables, and test the reliability of the results through repeated cross validation. The use of ML to determine a later outcome is known as forecasting. Forecasting is increasingly used in developing personalized medicine (e.g., using tissue biomarkers to predict the course of malignancies (Cruz & Wishart, 2006). Recent ML neuroimaging studies have shown promising results in predicting the course of neuropsychiatric disorders (Orrù, Pettersson-Yeo, Marquand, Sartori, & Mechelli, 2012).

Psychiatry currently relies on descriptive information. Establishing the usefulness and limits of non-invasive, low cost information such as this is an important baseline to build upon with new, potentially more invasive and more expensive methods for identifying risk. However, the implementation of ML methods to clinical observations is limited. Investigators have used clinical information to forecast violent behavior among outpatients with schizophrenia (Tzeng, Lin, & Hsieh, 2004) achieving moderate success of 76.2% positive prediction of later violent behavior. In contrast, ML has failed in forecasting the course of bipolar disorder from clinical data (Moore, Little, McSharry, Geddes, & Goodwin, 2012). A review of studies forecasting the risk for psychotic disorders (Strobl, Eack, Swaminathan, & Visweswaran, 2012) showed an advantage for ML-based Support Vector Machines (SVMs) classification. In this example, predictive accuracy ranged from 100%–78% relative to GLM-based variance accounted for ranging from 81%–67%. This study further demonstrated that forecasting from multimodal information (e.g., quality of life and neuroimaging data) increases the accuracy of prediction. This conclusion was also reached by Marini et al. (2007) who used ML random forests to compare the classification of PTSD based on the clinical assessment of PTSD alone versus the clinical assessment of PTSD along with other clinical assessment tools. This study demonstrated large improvements

(from predictive accuracy of 70.59% based only on PTSD symptom assessment to 78.43% integrating other symptoms).

The current work evaluates the use of ML to forecast chronic PTSD from data available shortly after a traumatic event. This study utilizes data from a previously published, fifteen months long, longitudinal study of 975 trauma-exposed individuals admitted to a general hospital emergency department within hours of their traumatic events (Shalev et al., 2012). In the current work, we compare a) forecasting accuracy based on all available information utilized indiscriminately, b) forecasting accuracy based on a subset of variables that are selected using a feature selection algorithm, and c) forecasting accuracy from ASD symptoms alone. The current study also compares the prediction of PTSD diagnosis at fifteen months with that of an empirically derived non-remitting PTSD symptoms trajectory (Galatzer-Levy et al., 2013). To meet the specific challenge of early prediction, this work uses data obtained within ten days of a traumatic event.

2. Materials and Methods

2.1. Participants and Procedures

Data for the current study come for the Jerusalem Trauma Outreach and Prevention Study (J-TOPS; (Shalev et al., 2012), [ClinicalTrial.Gov](https://clinicaltrials.gov/ct2/show/study/NCT00146900) identifier: NCT00146900). The J-TOPS combined a systematic outreach and comprehensive follow-up design with an embedded, randomized, controlled trial of early interventions. Sampling procedures and population parameters of study subjects are fully described in (Shalev et al., 2012).

Subjects in the current study were adults who were admitted to Hadassah University Hospital emergency department (ED) immediately following potentially traumatic events (age 18–70). Following identification in the ED, potential subjects were screened using a short telephone interview 9.21±3.20 days following ED admission. Those with acute PTSD symptoms were invited for clinical interviews, which took place 29.51±4.93 days after ED admission. Participants were re-evaluated five, nine, and fifteen months after ED admission regardless and blind of their participation in the nested clinical trial. Participants provided informed consent for all aspects of the study with procedures approved and monitored by the Hadassah University Hospital's institutional review board (IRB).

2.1.1. Current Study Sample—Included in this study are participants with valid data at 10 days post-trauma and at least two additional time points. The resulting sample consisted of 957 participants. The initial traumatic event exposure included motor vehicle accidents (84.1%), terrorist attacks (9.4%), work accidents (4.4%), and other incidents (2.0%). Participants in this study did not differ from the J-TOPS larger sample in gender distribution, age, ten days symptom severity, and the number of new traumatic events occurring during the study (for full description, see (Galatzer-Levy et al., 2013); Table 1).

2.2. Instruments

To forecast PTSD, we used *all information items* collected during participants' ED admissions and phone interviews during the first ten days following trauma. The resulting 68 items (alias, "features") include demographic data, ED observations, and instruments

administered at ten days. We considered both *individual items* and total psychometrics scores (see data preparation below) as valid initial features.

Event and ED features include type of traumatic event (motor vehicle accident/work accident/ terrorist attack/other incident), age, gender, ED blood pressure, ED pulse, self-reported ED pain level, prescribed opiates, non-opiate analgesics and anti-inflammatory agents, and documented head injury, loss of consciousness or whiplash injury, and time spent in the ED (Table 1).

Telephone interviews features include DSM IV PTSD symptoms as per the PTSD Symptom Scale (PSS) interviewer version (PSS-I; (Foa & Tolin, 2000) and additional Acute Stress Disorder symptoms per the Acute Stress Disorder Scale (ASDS) (Bryant, Moulds, & Guthrie, 2000). Other clinical information collected at this time point included were The Kessler-6 (K6), a 6-item self-report of depression and general distress (Kessler et al., 2002); interviewers' and participants' Clinical Global Impression (GSI) of severity (Guy, 1976), a four item post-traumatic cognition instrument summarizing the Posttraumatic Cognitions Inventory (PTCI (Foa, 1999)) four dimensions (counting on others, counting on oneself, dangerousness of the world, and self-blame). Survivors were additionally asked if they felt that they needed help (yes/no), how they perceived their current social support, and how well they perceived their behavior during the traumatic event. They also completed a four item coping efficacy screening instrument that evaluated their current capacity for (a) sustained task performance ["work functioning"], (b) interpersonal relations ["relationship functioning"], emotional regulation, and negative self-perception ["worthlessness"].

Missing data was minimal (0–7%) for the majority of items. Items with higher proportion of missing variables included work functioning (48.5% missing), head injury (40.0% missing), loss of consciousness (26.1%), ED pain level (51.0%), and duration of ED admission (13.3%).

3. Modeling Approach

3.1. Outcome Measures

Following a previous Latent Growth Mixture Modeling (LGMM) analysis of these data (Galatzer-Levy et al., 2013), the study's primary outcome measure was membership of a nonremitting PTSD symptom trajectory throughout the study's fifteen months determined by the posterior probability of class membership derived using LGMM (see figure 1). The nonremitting symptom trajectory included 17% of the study sample, accounted for the majority (71%) of PTSD cases at fifteen months, included cases with higher symptom severity, and was unaffected by cognitive behavioral therapy (CBT) received during the study by n=125 participants. Membership in the non-remitting trajectory is compared with membership in progressively remitting trajectories.

As an alternative outcome measure, and to meet the literature's frequent reference to PTSD as an outcome, we separately tested the prediction of DSM-IV PTSD diagnostic status at fifteen months.

3.2. Data Preparation

Prior to analysis, categorical variables were dummy coded and all continuous variables were normalized to have a range from 0–1 to reduce noise due to differences in scaling. Missing data was handled using the `knimpute` command in MATLAB. To perform imputation in data sets with missing values, we applied a non-parametric nearest neighbor method (Batista & Monard, 2003). Specifically, this method imputes each missing value of a variable with the present value of the same variable in the most similar instance according to Euclidian distance metric.

3.3 Machine Learning

Machine Learning in this work involved multiple recursive steps that together informed the identification of potentially predictive features and the evaluation of the predictive accuracy of selected features using classification algorithms.

To examine different ways of optimizing an ML-based approach to the particular case of forecasting PTSD we repeated these steps using different predictive features and outcome definitions. The reliability of each choice was tested using cross-validation.

For clarity, we describe the study ML procedures in a stepwise manner, while in reality its procedures were iterative and somewhat parallel. Result section follows a similar ‘didactic’ order.

3.3.1. Feature selection—We applied a Markov Boundary Induction algorithm for Generalized Local Learning (GLL) (Aliferis, Statnikov, Tsamardinos, Mani, & Koutsoukos, 2010) to identify sets of features that provided the most direct predictive power of our target variables. This approach initially identifies variables that demonstrate a univariate association to the target variable and removes all others. It then tests each retained variable’s association with the target variable while controlling for other retained variables and excludes those that become non-significant when controlling for other variables. The resulting list consists of predictors that are independently associated with the target. We used an algorithm from the Causal Explorer library (Statnikov, Tsamardinos, Brown, & Aliferis, 2010) with Fisher’s Z test, $p < .05$, and parameter *max-k* set to 1.

3.3.2. Classification algorithms—In the current analysis we utilize Linear Support Vector Machines (SVMs) with the *C* parameter =1. SVMs identify a linear hyperplane in high dimensional space (each predictor variable is a dimension) that accurately separates the sample in two known classes (Statnikov, Aliferis, Hardin, & Guyon, 2011) [in the current study ‘classes’ are a) membership in remitting vs. non-remitting trajectories or b) PTSD diagnostic status].

We compare the performance of Linear SVMs to 6 other ‘best in class’ classification algorithms (Statnikov, Aliferis, Hardin, et al., 2011) because we do not have an *a priori* reason to believe that SVMs are optimal in this context. Specifically, we compared a) Linear SVMs with the *C* parameter = 1 to: b) Linear SVMs with the *C* parameter optimized based on the current data, (c) Polynomial SVMs with *C* parameter optimized (Statnikov, Aliferis, Hardin, et al., 2011), (d) Random Forests which construct a multitude of decisions trees and

select the class based on the mode of the classes outputted by the individual trees (Breiman, 2001), (e) AdaBoost which is a meta-algorithm which is utilized in conjunction with other machine learning algorithms to improve classification performance (Freund & Schapire, 1995), (f) Kernel Ridge Regression with Radial basis function kernel (KRR-RBF Kernel) (Hastie et al., 2003), and (g) Bayesian Binary Regression (BBR) where coefficients are estimated based on three choices of probability distribution including Gaussian, exponential, or Laplace distributions (Genkin, Lewis, & Madigan, 2007). These algorithms were chosen for comparison because they represent current *best in class* machine learning classification methods (Statnikov, Aliferis, & Hardin, 2011).

3.3.3. Evaluating different configuration of predictive features—To evaluate whether the above feature selection maintains the predictive accuracy despite removing redundant items we compared case classifications made using selected features with those obtained using all available information items. We also compared the use of selected features with that of retaining only highly selected features (those that appear in 75% and 95% of cross-validation runs), and the use of feature selection with that of restricting the predictors to *PTSD and ASD symptoms*.

3.3.4. Comparison of outcome definitions—To compare different outcome definitions, we compared accuracy of predicting the non-remitting symptom trajectory, as defined above, and predicting PTSD status at 15 months.

3.3.5. Directionality of effects—To document the direction of the relationships between features selected and target outcome we conducted a series of t-tests and chi-square analyses of features that appear in over 75% of cross-validation runs.

3.3.6. Cross-validation—To avoid overfitting solutions and ensure the generalizability of our findings, all feature selection and classification experiments were conducted in a ten times repeated 10-folds cross-validation. In the 10-folds cross-validation, participants are randomly assigned into ten non-overlapping subsets containing approximately the same number of cases and non-cases. The classification algorithm is trained in nine of these ten data subsets, and subsequently (and independently) tested in the remaining tenth subset. This procedure is repeated iteratively, resulting in all tenths of the data being used for both training and testing of the algorithm. The entire procedure is repeated 10 times, resulting in a total of 100 runs (ten repetitions of ten trainings and testing). In the current study, a cross-validation algorithm was written in MATLAB version 8.3 (The Mathworks, 2012) to randomize cases into 90%/10% splits. The first 90% of the data was utilized in the feature selection algorithm and the best solution in that 90% was tested in the hold out 10%. This procedure was repeated 10 times. Features that were selected and confirmed across random splits of the sample were then introduced into the ML classification algorithms. Once again, data was randomly split using the same cross-validation procedure. For each classification algorithm, the solution was identified in a random 90% of the data and validated in the hold out 10%. This procedure was also repeated 10 times. Appendix One provides a full MATLAB script to conduct feature selection, cross-validation of feature selection, SVMs, cross-validation of SVMs, along with syntax to call the mean AUC across cross-validations

of the SVMs. The cross-validation procedure is on top of, rather than embedded in, the feature selection and classification algorithms.

3.3.7 Accuracy metrics—Estimates of predictive accuracy are expressed as Area Under the Receiver Operating Characteristics curve (AUC). The ROC curve is a plot of the sensitivity versus 1-specificity of a classification system, and infers the accuracy of that system, thereby creating a comparable metric across experiments. (Bradley, 1997). Following literature standards (Fawcett, 2003; Harrell, Lee, & Mark, 1996), we consider ROC curve AUC of .50–.60 as indicating prediction at chance; .60–.70 as indicating poor prediction; .70–.80 fair prediction; .80–.90 good prediction; .90–1.0 excellent prediction.

3.3.8. Total Machine Learning Feature Selection and Forecasting Algorithm—In total (following Figure 2), normalized data is utilized to identify the minimal set of features that provides the maximum unique information about the target variable (in this case trajectory membership) using a Markov Boundary feature selection algorithm (Step 1 in Figure 2). The results are established and confirmed using 10x10 fold cross-validation (Step 2). The selected (and validated) features are then utilized in the classification algorithms (Step 3). The results of the classification algorithms are also tested and confirmed using the same cross-validation method as the feature selection algorithm with an additional step used when utilizing algorithms that include additional optimized parameters (Step 4). The result is a mean AUC across 100 cross-validations of the classification algorithms based on feature selected variables (Step 5).

4. Results

4.1. Feature selection

The feature selection algorithm identified a parsimonious set of predictive features for the non-remitting PTSD symptom trajectory. Twenty-one features appeared in 75% of cross-validation runs, and sixteen in 95% of cross validation runs (Figure 2).

4.2. Comparing classification algorithms

There were no significant differences between classification algorithms when applied to selected features as indicated by completely overlapping C.I.s across AUCs for the different classification algorithms. The respective mean AUCs across 100 cross-validation runs were: Linear SVMs ($C = 1$) = .77 [95% CI=0.74–0.82]; Linear SVM (Optimized C) = .76 [95% CI=0.73–0.82]; Polynomial SVM (Optimized C) = .77 [95% CI=0.73–0.81]; Random Forests = 0.80 [95% CI=0.76–0.84]; AdaBoost = .77 [95% CI=0.75–0.83]; KRR with RBF Kernel = .79 [95% CI=0.75–0.83]; BBR = .77 [95% CI=0.74–0.82]. Because SVMs are well suited for data with heterogeneous non-normal distributions consistent with the kinds of data (i.e. self-report, biological) (Statnikov, Aliferis, Hardin, et al., 2011) that PTSD researchers are interested in integrating for prediction, for instructive purposes, we proceeded only with linear SVMs ($C=1$) for all subsequent evaluation of classification accuracy.

4.3. Comparing predictive features configurations

Linear SVMs ($C=1$) used to compare different feature configuration yielded the following: (a) for all variables: $AUC = .78$ [95% $CI=0.74-0.82$]; (b) for variables defined by feature selection $AUC = .77$ [95% $CI=0.73-0.81$]; (c) for features appearing in 75% of cross-validation runs, $AUC = .81$ [95% $CI=0.77-0.85$] and (d) for features appearing in 95% of cross-validation runs, $AUC = .82$ [95% $CI=0.78-0.86$]. Results were statistically equivalent as indicated by completely overlapping C.I.s across AUCs for the different classification algorithms.

4.4. Predicting from ASD symptoms

SVM-based prediction of symptom trajectories from ten days' ASD symptoms resulted in mean $AUC=.60$ [95% $CI=0.56-0.64$]; an AUC that falls into the range of 'prediction by chance'.

4.5. Predicting PTSD status

The prediction of PTSD diagnostic status yielded an $AUC = .71$ [95% $CI=0.67-0.75$], which is significantly weaker than the previously described $AUC=0.78$ [95% $CI=0.74-0.82$] obtained for PTSD symptom trajectories.

4.6. Directionality of relationship between predictors and outcome (Table 2)

For information obtained in the ED, higher age, being in a terrorist attack, having spent more time in the ED, experiencing more pain, having received tranquilizers, and reporting head injury or loss of consciousness were positively associated with the non-remitting trajectory, while being in a motor vehicle accident was positively associated with the remitting trajectory.

For information obtained at the 10-days interview, higher levels of ASD and PTSD and individual symptoms (nightmares, concentration difficulties, increased startle, flashbacks, irritability, derealization), as well as worthlessness, higher overall PTSD symptom severity, PTSD diagnosis, requesting help, higher patient and clinician CGIs, and lower levels of social support were all positively associated with the non-remitting PTSD trajectory (see Table 2).

5. Discussion

The current study evaluated the implementation of ML methods to forecast PTSD from information obtained within ten days of a traumatic event. The ML feature selection identified features that exhausted the predictive potential in the dataset. Of the latter, 16 were present in over 95% of repeated cross validations, yielding an AUC of 0.82. These features included, side-by-side, event and injury characteristics, ED parameters, psychometric measures and subjective self-appraisal. Selected features forecasting accuracy was equivalent to that obtained by using all features. Membership in a non-remitting PTSD symptom trajectory was more accurately predicted than PTSD status at 15 months. Predicting from ASD symptoms was worse than predicting from ML-selected features. The seven different classification algorithms yielded equal classification accuracies.

The finding that readily collected initial information, including ED data and commonly used, short assessment tools, provided consistent and efficient forecasting is very promising in that (a) such measures can be readily translated to clinically relevant prediction tools and (b) the cost of procuring them is minimal. The forecasting accuracy in this study is consistent with AUCs obtained in other psychiatric forecasting and classification studies (.81 and .76); (Marini et al., 2007; Strobl et al., 2012).

PTSD studies that examined theoretically derived predictors have identified promising candidates. However, replications of these findings were inconsistent [for a review see (Bryant, 2003)]. While many of the retrieved predictors have significant theoretical and heuristic value, their limited consistency may be due to inherent limitations of their modeling approaches. First, studies have focused on predicting PTSD diagnostic status. However, the PTSD outcome encompasses a wide range of symptom configurations and severities and may be too broad of a definition for focused modeling approaches (Galatzer-Levy & Bryant, 2013). Next, studies to date have utilized a GLM framework that, while useful for testing focused hypotheses, provide an average solution rather than seeking a consistent signal across subjects. Further, collinearity between closely related variables that share a common cause may obfuscate the identification of potentially predictive variables (Spirites, Glymour, & Scheines, 2000). The GLM framework, therefore, may be limited in its ability to match the disorder's multicausal etiology and the heterogeneous outcome. The ML approach presented here may be better equipped, since both the feature selection and the SVM classification do not rely on analysis of variance to identify relationships, and as such can determine probabilistic relationships across large sets of variables without errors due to collinearities or limited variable-to-sample ratios. Further, this approach has a built-in validation method to determine if models are accurate across subjects rather than on the aggregate (Aliferis et al., 2010; Statnikov, Aliferis, & Hardin, 2011).

ML's identification of features that may be highly collinear allows for the disaggregation of measures to identify specific components that signal risk. In the current study, we find that specific PTSD, ASD, and depression symptoms, despite being closely related, provided additional information. Beyond symptom levels, type of trauma, pain level, length of ED treatment, head injury, or loss of consciousness are independent and consistent risk indicators. This is consistent with other studies that have used ML classification and feature selection methods to identify clinical self-report characteristics that most accurately differentiate PTSD cases from non-cases. For example, (Marini et al., 2007) utilized Random Forests to demonstrate that clinical self-report measures not directly associated with PTSD, such as positive and negative self-reported symptom measures of other disorders, provide significant additive classification accuracy over PTSD symptom assessments on their own. ML methods, in this case SVMs, have also been used to build classifiers differentiating trauma survivors with PTSD from controls based on grey matter density measured using MRI (Gong et al., 2014). Neither of these studies attempted to forecast the course of PTSD based on acute features, however, making the current work novel in this way. However, the findings in the current study should be tested for their accuracy to forecast as well as classify posttraumatic stress outcomes. ML classification algorithms have also been used to forecast the need for life-saving interventions among trauma patients based on vital signs and heart-rate while in transport to the hospital with

very high accuracy (Liu, Holcomb, Wade, Darrah, & Salinas, 2014). This provides a useful example of how such methods can be used for efficient and accurate clinical decision-making.

The *marginal prediction from ASD symptoms* (which consist of all PTSD symptoms plus specific ASD-related symptoms) is consistent with previous findings showing that ASD, whilst identifying a high risk group, does not provide a robust predictive signal because most subjects who develop PTSD do not meet early ASD diagnostic criteria (Bryant, 2003). In this study, the failure to predict from ASD symptoms may additionally reflect the previously reported high frequency of these symptoms shortly after trauma exposure (Rothbaum, Foa, Riggs, Murdock, & Walsh, 1992). Predicting from frequently endorsed features yields highly sensitive but non-specific prediction (Shalev, Freedman, Peri, Brandes, & Sahar, 1997).

Additionally, initial PTSD symptoms are only one of several expressions of distress following traumatic events, and thus, including other responses in a predictive model (e.g., ED pain, expressing a need for help, sense of worthlessness; Figure 3) captures a wider array of initial responses. The inclusion of event- and injury characteristics (e.g., type of accident, head injury, length of stay in ED) may similarly add to the model's predictive power.

The relative *advantage of predicting the non-remitting PTSD symptom trajectory* may reflect the fact that this trajectory provides a more accurate description of the data, whereas survivors defined by PTSD status have a range of PTSD symptom severity. As shown in a recent longitudinal study (Bryant, O'Donnell, Creamer, McFarlane, & Silove, 2013), survivors fluctuate between formal PTSD diagnostic status, while showing limited change in symptom severity scores. Thus, the dichotomous PTSD yes/no outcome, despite providing a formally desirable frame of reference, results in loss of information and, as this work shows, lesser predictability. Additionally, the non-remitting trajectory in this sample was also treatment refractory, and thus insensitive to the potential confounding effect of receiving treatment. Our results tentatively suggest that outcome variables that capture inherent group heterogeneities may be a better target for prediction.

We used SVMs as the primary approach because this algorithm works well with relatively small sample sizes and a high number of features. SVM can also accommodate both simple and complicated models, and has strong built-in mechanisms to avoid over-fitting (Statnikov, Aliferis, Hardin, et al., 2011). The consistency of AUCs levels across ML classification algorithms suggests that this work has exhausted the predictive information in the data attainable through existing ML methods.

Limitations to this work include the lack of information about other known predictors [e.g., prior psychiatric history, peri-traumatic responses, negative cognitions (Ozer, Best, Lipsey, & Weiss, 2000)]. Additionally, this work does not include biological or neurocognitive measures (Bryant, 2003). Including such variables in future studies may improve the prediction of PTSD.

This work is also limited by addressing PTSD and PTSD symptoms as sole outcome dimension, whereas traumatic stress can lead to other mental disorders and conditions (e.g., depression, substance abuse). Studies predicting cognitive, vocational, and biological outcomes may address other dimensions of the response to traumatic events. Finally, by only including an ED-based sample (as opposed to protracted traumatization, e.g., during wars) this work is also limited to exploring one of several pathways to chronic PTSD. Specifically, it does not explore delayed-onset PTSD (deRoon-Cassini, Mancini, Rusch, & Bonanno, 2010).

The current study is far from offering an optimal or comprehensive solution to predicting PTSD. Nonetheless it demonstrates the feasibility and the yield of ML-based prediction from early responses to traumatic events. It thereby provides a necessary foundation for extending the use of an ML approach to forecasting PTSD in other studies and to including biomarkers in the model. While the current study attempted to validate its findings using cross-validation methods, truly usable methods for identifying individuals at risk require replication and improvement in independent samples. An advantage of using a learning framework is that additional sources of information can be integrated and the accuracy and generalizability of forecasting can be improved upon. As such, the current work provides a baseline indicating the conjoined predictive accuracy of well-known or commonly occurring assessments conducted in ED setting and among individuals thought to be at risk for developing posttraumatic stress pathology. These results should be validated in independent datasets and improved upon with additional sources of information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by a research grant MH071651 from the National Institute of Mental Health. The study [ClinicalTrial.gov](https://clinicaltrials.gov/ct2/show/study/NCT00146900) identifier is NCT00146900.

We acknowledge members of the J-TOPS group: Yael Ankri, M.A.; Sara Freedman, Ph.D.; Rhonda Addesky, Ph.D.; Yossi Israeli – Shalev, M.A.; Moran Gilad, M.A.; Pablo Roitman M.D.

Appendices

Appendix I: MATLAB syntax for calling feature selection and classification algorithms.

References

- Aliferis CF, Statnikov A, Tsamardinos I, Mani S, Koutsoukos XD. Local Causal and Markov Blanket Induction for Causal Discovery and Feature Selection for Classification Part I: Algorithms and Empirical Evaluation. *J. Mach. Learn. Res.* 2010; 11:171–234.
- Batista GE, Monard MC. An analysis of four missing data treatment methods for supervised learning. *Applied Artificial Intelligence.* 2003; 17(5–6):519–533.
- Boscarino JA, Erlich PM, Hoffman SN, Zhang X. Higher FKBP5, COMT, CHRNA5, and CRHR1 allele burdens are associated with PTSD and interact with trauma exposure: implications for neuropsychiatric research and treatment. *Neuropsychiatric disease and treatment.* 2012; 8:131–139. [PubMed: 22536069]

- Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*. 1997; 30(7):15.
- Breiman L. Random Forests. *Machine Learning*. 2001; 45(1):5–32.
- Brewin CR, Andrews B, Valentine JD. Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. *J Consult Clin Psychol*. 2000; 68(5):748–766. [PubMed: 11068961]
- Bryan CJ, Clemans TA. Repetitive traumatic brain injury, psychological symptoms, and suicide risk in a clinical sample of deployed military personnel. *JAMA Psychiatry*. 2013; 70(7):686–691. [PubMed: 23676987]
- Bryant RA. Early predictors of posttraumatic stress disorder. *Biological psychiatry*. 2003; 53(9):789–795. [PubMed: 12725971]
- Bryant RA, Moulds ML, Guthrie RM. Acute Stress Disorder Scale: a self-report measure of acute stress disorder. *Psychological Assessment*. 2000; 12(1):61–68. [PubMed: 10752364]
- Bryant RA, O'Donnell ML, Creamer M, McFarlane AC, Silove D. A Multisite Analysis of the Fluctuating Course of Posttraumatic Stress Disorder Analysis of PTSD. *JAMA Psychiatry*. 2013:1–8.
- Cohen J. The earth is round ($p < .05$). *American Psychologist*. 1994; 49(12):997–1003.
- Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer informatics*. 2006; 2:59. [PubMed: 19458758]
- deRoon-Cassini TA, Mancini AD, Rusch MD, Bonanno GA. Psychopathology and resilience following traumatic injury: A latent growth mixture model analysis. *Rehabilitation Psychology*. 2010; 55(1):1–11. [PubMed: 20175629]
- Etkin A, Wager TD. Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *The American Journal of Psychiatry*. 2007; 164(10):1476. [PubMed: 17898336]
- Fawcett T. ROC Graphs: Notes and Practical Considerations for Researchers. Technical Report, HPL-2003-4, HP Laboratories. 2003
- Foa EB, Ehlers A, Clark DM, Tolin DF, Orsillo SM. The Posttraumatic Cognitions Inventory (PTCI): Development and Validation. *Psychological Assessment*. 1999; 11:303–314.
- Foa EB, Tolin DF. Comparison of the PTSD Symptom Scale–Interview Version and the Clinician-Administered PTSD Scale. *Journal of Traumatic Stress*. 2000; 13(2):181–191. [PubMed: 10838669]
- Freund, Y.; Schapire, RE. A decision-theoretic generalization of on-line learning and an application to boosting; Paper presented at the Computational learning theory; 1995.
- Galatzer-Levy IR, Ankri Y, Freedman S, Israeli-Shalev Y, Roitman P, Gilad M, Shalev AY. Early PTSD Symptom Trajectories: Persistence, Recovery, and Response to Treatment: Results from the Jerusalem Trauma Outreach and Prevention Study (J-TOPS). *PLoS one*. 2013; 8(8):e70084. [PubMed: 23990895]
- Galatzer-Levy IR, Bryant RA. 636,120 Ways to Have Posttraumatic Stress Disorder. *Perspectives on Psychological Science*. 2013; 8(6):651–662.
- Genkin A, Lewis DD, Madigan D. Large-scale Bayesian logistic regression for text categorization. *Technometrics*. 2007; 49(3):291–304.
- Gong Q, Li L, Tognin S, Wu Q, Pettersson-Yeo W, Lui S, Mechelli A. Using structural neuroanatomy to identify trauma survivors with and without post-traumatic stress disorder at the individual level. *Psychological medicine*. 2014; 44(01):195–203. [PubMed: 23551879]
- Guy W. Clinical global impression scale. The ECDEU Assessment Manual for Psychopharmacology-Revised. 1976; 338:218–222. Volume DHEW Publ No ADM 76, 338.
- Hald, A. A History of Parametric Statistical Inference from Bernoulli to Fischer. Springer; 2007. p. 1713-1935.
- Harrell F, Lee KL, Mark DB. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*. 1996; 15:361–387. [PubMed: 8668867]
- Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer; 2003.

- Karl A, Schaefer M, Malta LS, Dorfel D, Rohleder N, Werner A. A meta-analysis of structural brain abnormalities in PTSD. *Neurosci Biobehav Rev.* 2006; 30(7):1004–1031. [PubMed: 16730374]
- Kessler RC. Posttraumatic stress disorder: the burden to the individual and to society. *J Clin Psychiatry.* 2000; 61(Suppl 5):4–12. discussion 13–14. [PubMed: 10761674]
- Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SLT, Zaslavsky AM. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine.* 2002; 32(6):959–976. [PubMed: 12214795]
- Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB. Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry.* 1995; 52:1048–1060. [PubMed: 7492257]
- Liu NT, Holcomb JB, Wade CE, Darrah MI, Salinas J. Utility of vital signs, heart-rate variability and complexity, and machine learning for identifying the need for life-saving interventions in trauma patients. *Shock (Augusta, Ga.).* 2014
- Marini I, Supek F, Kova i0107; Z, Rukavina L, Jendri ko T, Kozari Kova i D. Posttraumatic stress disorder: diagnostic data analysis by data mining methodology. *Croatian medical journal.* 2007; 48(2):185–197. [PubMed: 17436383]
- Moore P, Little M, McSharry P, Geddes J, Goodwin G. Forecasting Depression in Bipolar Disorder. 2012
- Orru G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neuroscience & Biobehavioral Reviews.* 2012; 36(4):1140–1152. [PubMed: 22305994]
- Ozer E, Best S, Lipsey T, Weiss D. Predictors of Posttraumatic Stress Disorder Symptoms in Adults: A meta-analysis. Manuscript submitted for publication. 2000
- Ozer EJ, Best SR, Lipsey TL, Weiss DL. Predictors of Posttraumatic Stress Disorder and Symptoms in Adults: A Meta-Analysis. *Psychological Bulletin.* 2003; 129(1):52–73. [PubMed: 12555794]
- Peleg T, Shalev AY. Longitudinal studies of PTSD: overview of findings and methods. *CNS spectrums.* 2006; 11(8):589. [PubMed: 16871125]
- Ressler KJ, Mercer KB, Bradley B, Jovanovic T, Mahan A, Kerley K, Myers AJ. Posttraumatic stress disorder is associated with PACAP and the PAC1 receptor. *Nature.* 2011; 470(7335):492–497. [PubMed: 21350482]
- Rothbaum BO, Foa EB, Riggs DS, Murdock T, Walsh W. A prospective examination of post-traumatic stress disorder in rape victims. *Journal of Traumatic Stress.* 1992; 5(3):455–475.
- Shalev AY, Ankri Y, Israeli-Shalev Y, Peleg T, Adessky R, Freedman S. Prevention of posttraumatic stress disorder by early treatment: results from the Jerusalem Trauma Outreach and Prevention Study. *Archives of general psychiatry.* 2012; 69(2):166. [PubMed: 21969418]
- Shalev AY, Freedman S, Peri T, Brandes D, Sahar T. Predicting PTSD in trauma survivors: prospective evaluation of self-report and clinician-administered instruments. *Br J Psychiatry.* 1997; 170:558–564. [PubMed: 9330024]
- Spirtes, P.; Glymour, C.; Scheines, R. Causation, prediction, and search. Vol. 81. The MIT Press; 2000.
- Statnikov, A.; Aliferis, CF.; Hardin, DP. A Gentle Introduction to Support Vector Machines in Biomedicine: Theory and Methods. Vol. 1. World Scientific Publishing Company; 2011.
- Statnikov, A.; Aliferis, CF.; Hardin, DP.; Guyon, I. A Gentle Introduction to Support Vector Machines in Biomedicine: Case Studies. World Scientific Publishing Co., Inc.; 2011.
- Statnikov A, Tsamardinos I, Brown LE, Aliferis CF. Causal explorer: A matlab library of algorithms for causal discovery and variable selection for classification. *Causation and Prediction Challenge Challenges in Machine Learning.* 2010; 2:267.
- Stigler, SM. The history of statistics: The measurement of uncertainty before. Vol. 1900. Harvard University Press; 1986.
- Strobl EV, Eack SM, Swaminathan V, Visweswaran S. Predicting the risk of psychosis onset: advances and prospects. *Early Intervention in Psychiatry.* 2012; 6(4):368–379. [PubMed: 22776068]
- The Mathworks I. MATLAB and Statistics Toolbox (Version Release 2012b). Massachusetts, United States: 2012.

Tzeng, H-m; Lin, Y-L.; Hsieh, J-G. Forecasting violent behaviors for schizophrenic outpatients using their disease insights: Development of a binary logistic regression model and a support vector model. *International Journal of Mental Health*. 2004; 33(2):17–31.

Research Highlights

1. Machine Learning (ML) forecasting improves prediction of PTSD course.
2. Empirical trajectories are more predictable than PTSD diagnosis.
3. ML feature selection improves identification of unique predictors of PTSD course.

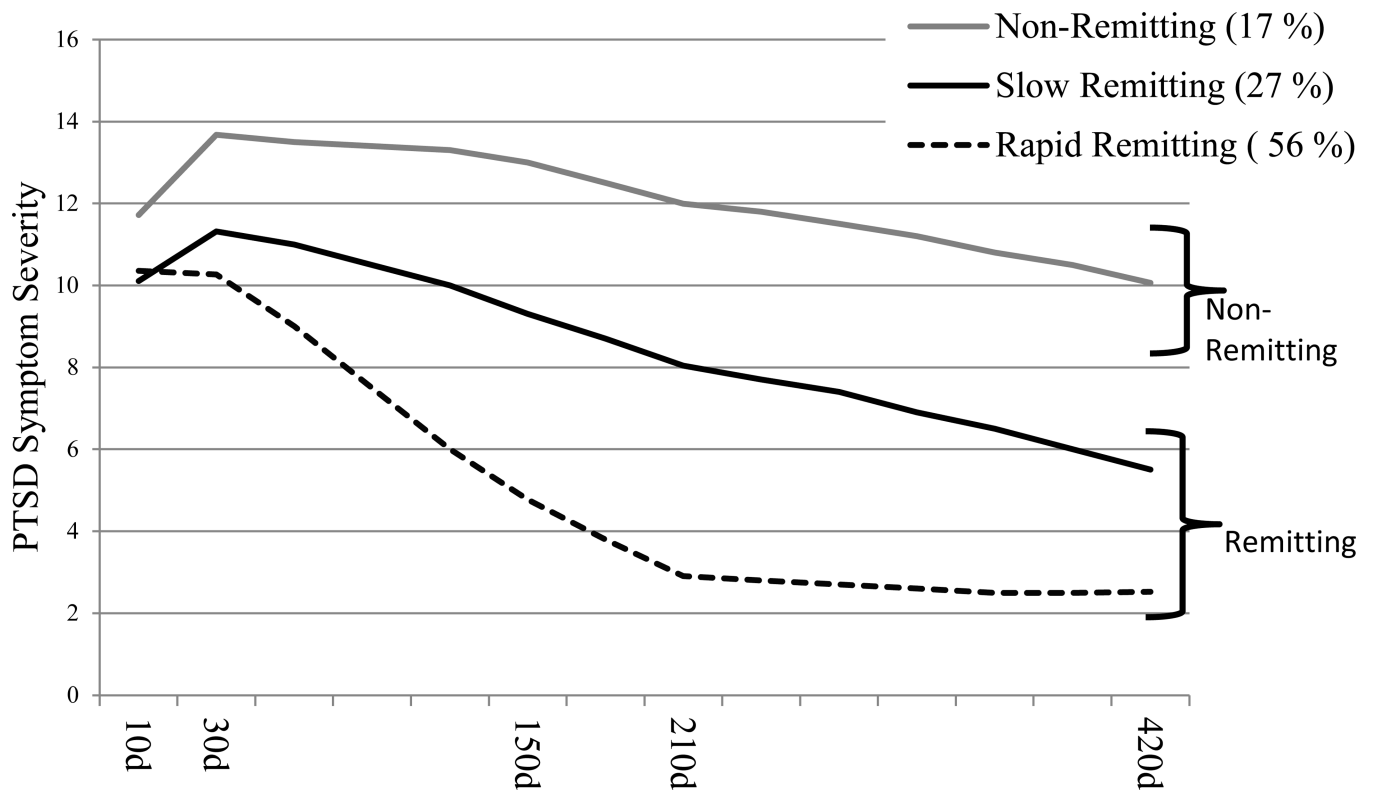


Figure 1. Three Trajectory Model of PTSD Symptom Severity Recovery Trajectories (n=957)

Note: x-axis indicates number of PTSD symptoms reported on the PSS-I. Y-axis represents time from ~10 days to ~15 months Trajectories represent estimated marginal means. 'd' indicates days from emergency room admission. Individuals are identified as members of modeled trajectories based on their posterior probability of class membership derived using Latent Growth Mixture Modeling.

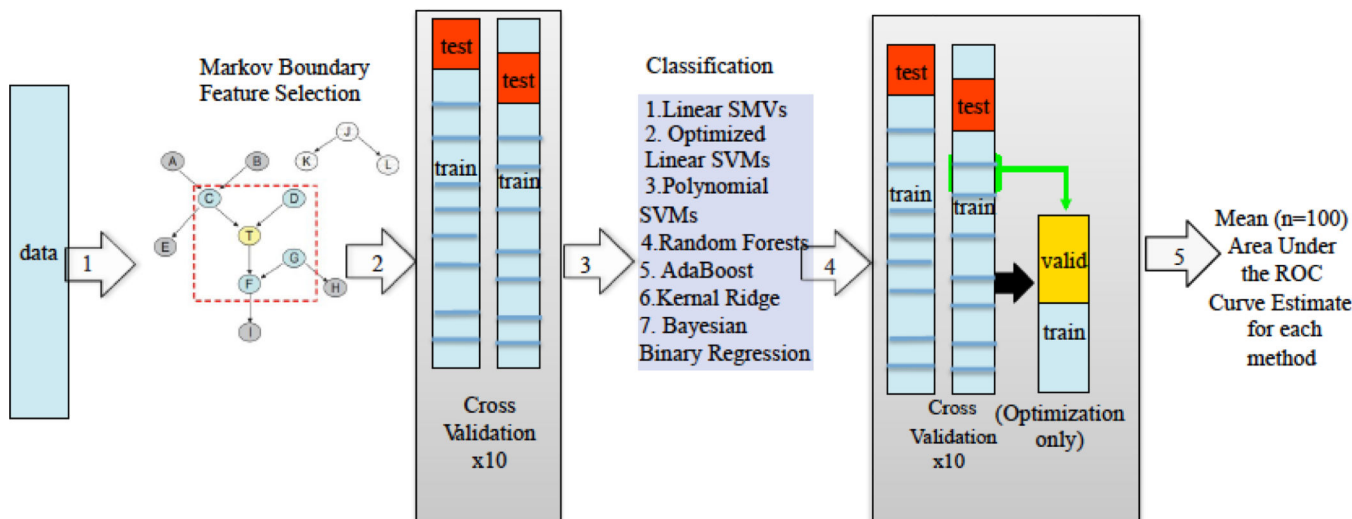


Figure 2. Machine Learning Approach for feature selection and classification

Note: 1. Unselected data are organized such that all potential predictors are normalized to ranges of 0–1 and a ‘target’ variable is specified; 2+3. The Markov Boundary Feature Selection algorithm removes redundant or uninformative variables to identify an irreducible set of predictors in a random 90% of cases. This set of predictors is then confirmed in a random 10% of cases, and the procedure is repeated 10 times; 4+5. Selected features are fed into seven different classification algorithms to determine the accuracy of selected features to classify the ‘target’ variable and to provide an accuracy estimate using area under the receive operator characteristic curve (AUC). The classification algorithms are tested using the same cross-validation procedure as for feature selection. Additionally, when optimizing parameters of the model (for example for the polynomial SVMs) an additional step of splitting each training set into a training & validation set is added; f. A mean AUC across 100 cross-validation runs is provided to determine the overall accuracy of the selected and validated features for classifying the target variable (in this analysis remission vs. non-remission trajectory membership or PTSD diagnostic status). SVM = Support Vector Machine; ROC = Receiver Operator Characteristic.

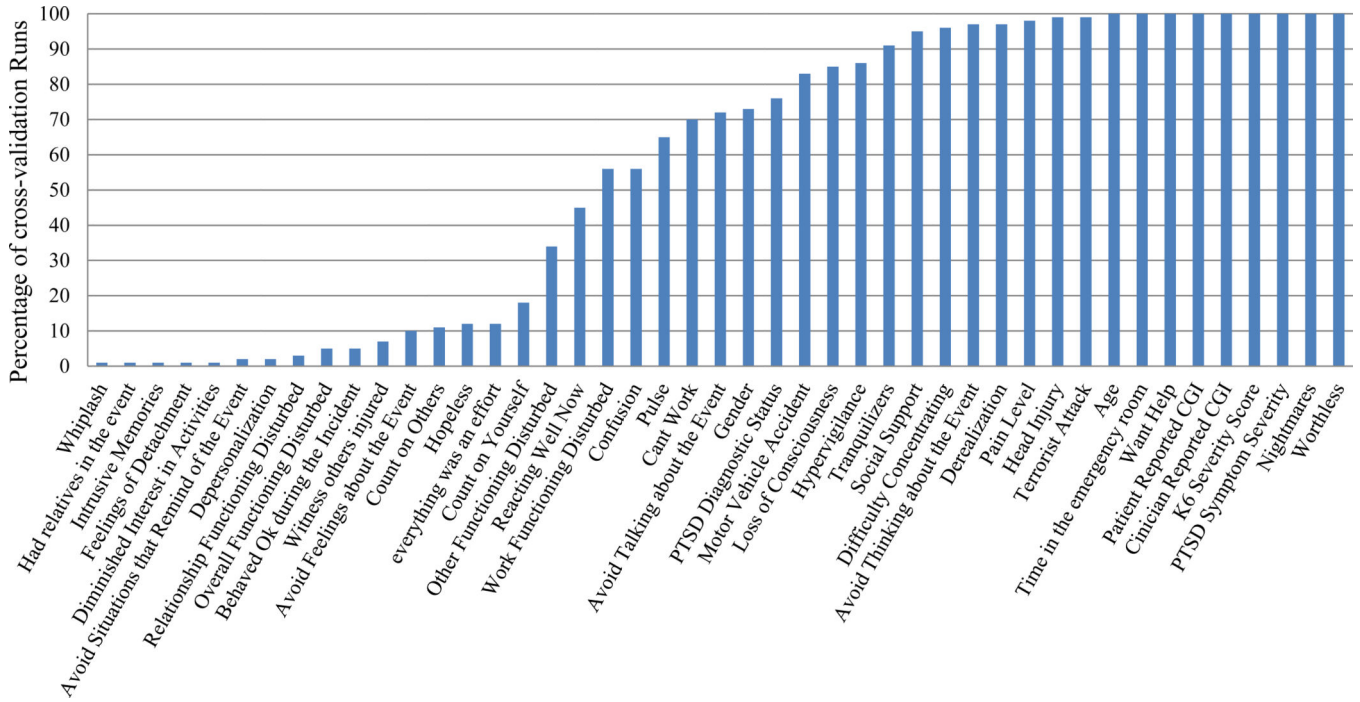


Figure 3. Features Selected Using Generalized Local Learning (GLL-MB) Algorithm across 100 cross validation runs

The figure is a graphical representation of the percentage of cross-validation runs for which each feature was selected. Features that were selected at a high frequency indicate that they consistently provide unique predictive information. CGI refers to the Clinical Global Impression metric.

Table 1

Predictive Features

	Feature Category	Features	Mean (SD)/%
Emergency Room	<i>Demographics</i>	Age	36.29(12.04)
		Time in ED (hours)	5.71 (6.52)
		Gender (% Male)	51.10%
	<i>Physiological responses</i>	Max Blood Pressure	129.99(17.25)
		Min Blood Pressure	78.53(11.42)
		Pulse	84.88(14.54)
		Pain Level (1-10 scale)	6.64 (2.58)
	<i>Pain Medications</i>	Opiates	8.80%
		non-opiate analgesics	36.20%
		Anti-inflammatory's	8.20%
	<i>Head Injury</i>	Head Injury	25.00%
		Whiplash	17.60%
		Loss of Consciousness	4.90%
	<i>Event Type</i>	Motor Vehicle Accident	85.10%
		Work Accident	4.20%
Terrorist Attack		9.00%	
10-Days Post Event	<i>Event Characteristics</i>	Witness others injured	29.80%
		Had relatives in the event	47.40%
	Coping efficacy	Overall Functioning	82.10%
		Work Functioning Disturbed	73.80%
		Relationship Functioning	40.70%
		Other Functioning Disturbed	63.30%
		Social Support	72.30%
	<i>Post-traumatic</i>	Count on Others (1-100 scale)	42.5 (29.4)
		Count on Yourself (1-100)	69.1 (29.6)
		Dangerous World (1-100)	71.2(27.4)
		Blame Yourself (1-100 scale)	22.4(32.1)
		Want Help	45.80%
	Perceived initial	Can't Work	26.00%
		Behaved Ok During the Event	79.50%
	<i>Clinical Assessments</i>	Reacting Well Now	77.10%
Patient Reported CGI (1-7)		4.29(1.57)	
Clinician Reported CGI (1-7)		4.11 (1.22)	
PTSD Symptom (1-17 scale)		10.27(3.24)	
K6 Severity Score (3-30)		17.8(5.18)	

Table 2

Means (SD) and frequencies for selected features appearing in 75 cross validation runs

	Feature	Non-Remitting Class	Remitting Classes
Emergency Room	Age	39.37(12.27)	35.58(11.74) ***
	Exposure to a Terrorist Attack	16.0%	7.6% ***
	Motor Vehicle Accident	79.8%	86.1% *
	Time in the Emergency Room	8.40(9.67)	5.20(5.60) ***
	Pain Level	7.12(2.33)	6.56(2.60) ⁺
	Tranquilizers	5.8%	1.6% **
	Head Injury	53.3%	39.0% **
	Loss of Consciousness	12.7%	5.3% **
10 days post-	Patient Reported CGI	5.12(1.43)	4.12(1.54) ***
	Clinician Reported CGI	4.83(1.27)	3.96(1.16) ***
	PTSD Symptom Severity	12.07(3.11)	9.94 (3.22) ***
	PTSD Diagnostic Status	86.5%	69.9% ***
	Nightmares	59.9%	33.5% ***
	Difficulty Concentrating	88.7%	64.6% ***
	Increased Startle	75.2%	56.0% ***
	Flashbacks	71.6%	54.6% ***
	Irritability	72.3%	55.4% ***
	Derealization	50.0%	33.35% ***
	Worthlessness	2.98(1.49)	2.05 (1.26) **
	Social Support	55.7%	75.6% ***
	Wanting Help	68.5%	41.1% ***

Note: All continuous variables were compared using independent samples T-Tests; all categorical variables were compared using Pearson's χ^2 statistic;

⁺ $p < .07$,

* $p < .05$,

** $p < .01$,

*** $p < .001$.