Posttraumatic Stress Disorder: Diagnostic Data Analysis by Data Mining Methodology

Igor Marinić[1], Fran Supek[2], Zrnka Kovačić[3], Lea Rukavina[4], Tihana Jendričko[1], Dragica Kozarić-Kovačić[1]

[1]Dubrava University Hospital, Department of Psychiatry, Referral Center of the Ministry of Health and Social Welfare for Stress-related Disorders, Zagreb, Croatia
[2]Laboratory for information systems, Department of Electronics, Ruđer Bošković Institute, Zagreb, Croatia
[3]Croatian Institute for Brain Research, University of Zagreb School of Medicine, Croatia
[4]"Bonifarm", Polyclinic for Clinical Pharmacology and Toxicology, Zagreb, Croatia

> **Correspondence to:**

Dragica Kozarić-Kovačić
Department of Psychiatry
Dubrava University Hospital
Avenija Gojka Šuška 6
10000 Zagreb, Croatia
dkozaric_kovacic@yahoo.com

> **Received:** December 6, 2006
> **Accepted:** March 8, 2007

> **Croat Med J. 2007;48:185-97**

Aim To use data mining methods in assessing diagnostic symptoms in posttraumatic stress disorder (PTSD).

Methods The study included 102 inpatients: 51 with a diagnosis of PTSD and 51 with psychiatric diagnoses other than PTSD. Several models for predicting diagnosis were built using the random forest classifier, one of the intelligent data analysis methods. The first prediction model was based on a structured psychiatric interview, the second on psychiatric scales (Clinician-administered PTSD Scale – CAPS, Positive and Negative Syndrome Scale – PANSS, Hamilton Anxiety Scale – HAMA, and Hamilton Depression Scale – HAMD), and the third on combined data from both sources. Additional models placing more weight on one of the classes (PTSD or non-PTSD) were trained, and prototypes representing subgroups in the classes constructed.

Results The first model was the most relevant for distinguishing PTSD diagnosis from comorbid diagnoses such as neurotic, stress-related, and somatoform disorders. The second model pointed out the scores obtained on the CAPS scale and additional PANSS scales, together with comorbid diagnoses of neurotic, stress-related, and somatoform disorders as most relevant. In the third model, psychiatric scales and the same group of comorbid diagnoses were found to be most relevant. Specialized models placing more weight on either the PTSD or non-PTSD class were able to better predict their targeted diagnoses at some expense of overall accuracy. Class subgroup prototypes mainly differed in values achieved on psychiatric scales and frequency of comorbid diagnoses.

Conclusion Our work demonstrated the applicability of data mining methods for the analysis of structured psychiatric data for PTSD. In all models, the group of comorbid diagnoses, including neurotic, stress-related, and somatoform disorders, surfaced as important. The important attributes of the data, based on the structured psychiatric interview, were the current symptoms and conditions such as presence and degree of disability, hospitalizations, and duration of military service during the war, while CAPS total scores, symptoms of increased arousal, and PANSS additional criteria scores were indicated as relevant from the psychiatric symptom scales.

Posttraumatic stress disorder (PTSD) is characterized by the symptoms of re-experiencing the traumatic event, avoidance symptoms, and increased arousal (1), but differences can be found in clinical presentations of symptoms between survivors of different traumas (2). Various comorbid diagnoses can be identified in these patients: alcohol abuse, depression, anxiety disorders, panic disorder and phobia, psychosomatic disorder, personality disorder, psychotic disorders, drug abuse, and dementia (3). Furthermore, PTSD is commonly misdiagnosed, resulting in inappropriate treatment (4).

Many authors reported various difficulties in estimating symptom severity in PTSD patients, especially if the diagnostic process was related to compensation seeking (3,5-8). It has also been shown that clinicians have a more subjective approach to patients who demand compensation (9).

Because of these adverse factors, the process of diagnosing PTSD is a complex one, and defining accurate diagnostic methods for PTSD is important in both clinical and forensic practice.

As some studies have shown (10,11), various data collected from the patient, such as short medical history, laboratory tests, specialist findings, or examination results could be analyzed with specialized "data mining" algorithms. Such algorithms can be used for finding intercorrelations between different parameters and exploring possibilities of using the acquired data for deriving rules and conditions useful in making faster diagnostic procedures and more targeted therapeutic interventions (12).

Data mining is defined as "nontrivial extraction of implicit, previously unknown, and potentially useful information from data" (13) and "the science of extracting useful information from large data sets or databases" (14). The term "data mining" includes both statistical techniques and algorithms developed for machine learning applications. Machine learning is a part of the artificial intelligence field in computer science, dealing with algorithms that can identify important structural features and seek non-obvious patterns in available data (15), in order to improve their performance in future, previously unseen situations. Especially relevant for data mining are those machine learning approaches which represent the acquired knowledge in an explicit form understandable to humans.

Since one of the major conditions for applying data mining techniques is the existence of uniform data sets (16), such methods are mostly used in biomedical research efforts, such as gene expression and regulation, protein structure, mutation research (17-19), and less frequently in everyday clinical work. To demonstrate that data mining techniques can assist in the diagnostic process of PTSD, we investigated the feasibility of a data mining approach applied to medical information acquired on psychiatric patients in the Department of Psychiatry of the Dubrava University Hospital, with the assistance of the experts from Ruđer Bošković Institute.

Subjects and methods

Subjects

The study included a total of 102 inpatients at the Department of Psychiatry, Dubrava University Hospital. One group comprised patients (veterans of 1991-1995 war in Croatia) who had a confirmed diagnosis of combat-related PTSD (n = 51), and the other group was a comparison group consisting of patients who had other psychiatric diagnoses but not PTSD (n = 51).

Patients were chosen by random number generator from 2450 patients treated at the Department of Psychiatry, Dubrava University Hospital, the Referral Center for Stress-related Disorders of the Ministry of Health and Social Welfare. A structured psychiatric interview and the Structured Clinical Interview for DSM-IV (20) (SCID) were applied to all patients. We also used Clinician-Administered PTSD Scale

(CAPS) (21), Positive and Negative Syndrome Scale (PANSS) (22), Hamilton Anxiety Scale (HAMA) (23), and Hamilton Depression Scale (HAMD) (24). The final diagnoses were determined by the psychiatrists, according to the International Classification of Disorders (ICD-10) (25) criteria. The patients were divided into two groups: patients with combat PTSD and patients with other psychiatric diagnoses (mood disorders, schizophrenia, schizotypal and delusional disorders, neurotic, stress-related, somatoform disorders, personality disorders). All patients completed the same diagnostic procedure.

All patients were men with median age (10th percentile-90th percentile of age distribution) of 38 (32.1-50.9). The groups were age matched ($P = 0.124$, two-tailed Mann-Whitney U-test), with the median age of PTSD patients equaling 38 (32.0-54.0) years and of non-PTSD patients 38 (33.0-45.0) years. Non-parametric statistics were used here due to age distribution in both groups differing from the normal distribution (Shapiro-Wilk test, P values were 0.013 and 0.014 for PTSD and non-PTSD groups, respectively).

Most of the PTSD patients were married (n = 39; 76.5%); 25.5% (n = 13) were without children, 17.6% (n = 9) had one child, 37.3% (n = 19) had two children, and 19.6% (n = 10) had three or more children. Regarding their education, 2.0% (n = 1) did not finish primary school, 7.8% (n = 4) had only elementary education, and 90.2% (n = 46) had secondary or higher education. Among the patients, 25.5% (n = 13) were employed, 7.8% (n = 4) unemployed, 27.5% (n = 14) were on sick leave, and 39.2% (n = 20) were retired.

In the group of non-PTSD patients, a similar percentage of patients was married (n = 38; 74.5%); 19.6% (n = 10) were without children, 25.5% (n = 13) had one child, 37.3% (n = 19) had two children and 17.7% (n = 9) had three or more children. As in the PTSD group, only 2.0% (n = 1) of patients did not finish primary school, while 5.9% (n = 3) had only elementary education, and 92.2% (n = 47) had secondary or higher education. Of all non-PTSD patients 21.6% (n = 11) were employed, 25.5% (n = 13) unemployed, 31.4% (n = 16) were on sick leave, and 21.6% (n = 11) were retired.

Methods

Each patient underwent the following examinations: 1) psychiatric (psychiatric interview and examination, clinical psychiatric scales, demographic data, evaluation of traumatic event, evaluation of health status and functioning); 2) psychological (neuropsychologic tests, personality tests, traumatic events questionnaires); and 3) biochemical examination (serum cortisol, thyroid hormones, lipids, serotonin, and dopamine indicators). Examination and collection of data was performed by a psychiatrist.

In order to be included in the study, patients had to be free of acute somatic disease, cardiocerebrovascular disease, diabetes or liver disease, and not abusing psychoactive substances or alcohol.

Only clinical data collected during the structured psychiatric interview, SCID, and the psychiatric scales were used in the analysis, including 80 of the 243 total attributes.

Diagnosis of PTSD and other diagnoses were made according to the ICD-10 criteria (25). The final diagnosis was made in cases where all sets of criteria (psychiatric and psychometric) were fulfilled. For all patients, the main diagnosis and first comorbid diagnosis, if available, were recorded.

To ensure a more uniform representation of comorbid diagnoses of PTSD patients in the database, some of the diagnoses were grouped, following the ICD-10 classification (25): A – no comorbid diagnosis (n = 15; 29.4% of PTSD patients); B – mood disorders (n = 12; 23.5%): bipolar affective disorder, depressive episode, recurrent depressive disorder, persistent affective disorders, other affective disorders; C – schizophrenia, schizotypal and delusional disorders (n = 8; 15.7%): schizophrenia, persistent delusional disorders, acute and transient psychotic disorders, unspecified nonorganic psychosis; D – neurotic, stress-related, and somatoform disorders (n = 13; 25.5%): other anxiety disorders, reaction to severe stress and adjustment disorders, other neurotic disorders; E – disorders of adult personality and behavior (n = 3; 5.9%).

The major diagnoses of non-PTSD patients were grouped in a similar manner: B – mood disorders (n = 10; 19.6% of non-PTSD patients): bipolar affective disorder, depressive episode, recurrent depressive disorder, persistent affective disorders, other affective disorders; C – schizophrenia, schizotypal and delusional disorders (n = 6; 11.8%): schizophrenia, persistent delusional disorders, acute and transient psychotic disorders, unspecified nonorganic psychosis; D – neurotic, stress-related, and somatoform disorders (n = 31; 60.8%): other anxiety disorders, reaction to severe stress and adjustment disorders, other neurotic disorders; E – disorders of adult personality and behavior (n = 4; 7.8%).

Comorbid diagnoses of non-PTSD patients were also grouped in the following way: B – mood disorders (n = 1; 2.0%): bipolar affective disorder, depressive episode, recurrent depressive disorder, persistent affective disorders, other affective disorders; D – neurotic, stress-related, and somatoform disorders (n = 4; 7.8%): other anxiety disorders, reaction to severe stress and adjustment disorders, other neurotic disorders; E – disorders of adult personality and behavior (n = 3; 5.9%).

Structured psychiatric interview collected data that included patient's age, education, qualification, employment, and marital status, and information about current relationships, number of children, housing and economic status, income, presence and degree of disability, military service, and legal offenses. It also included data on previous treatments, comorbid diagnoses, suicide attempts, previous and current social functioning, heredity, relations in the primary family, alcohol consumption, and smoking habits (5).

The CAPS scale is a structured interview that measures the frequency and intensity of main and associated symptoms of PTSD, and is one of the standard criteria for measuring traumatic stress (21). It consists of several components measuring the major clusters of PTSD symptoms. For each patient, the scores of traumatic event, re-experiencing symptoms, avoidance symptoms, and symptoms of hyperarousal, along with total CAPS scores and validity were entered in a database, while the two components that measure whether the duration of the disturbance exceeds one month and whether the disturbance causes clinically significant distress or impairment were not used because of highly unbalanced distributions (data not shown).

As both the scores and the validity are considered relevant for the assessment of a patient, for the purpose of this study we also calculated weighted scores (criterion score multiplied by [5 – validity score]) for each patient. By multiplying these values, the score and the validity both contributed to the final (weighted) score. Our intent was to investigate whether such scores will be identified as more important by the data mining algorithm than non-weighted scores.

The PANSS scale (22) was used to evaluate positive and negative syndromes in schizophrenia, and their association with global psychopathology. PANSS total positive scale scores, total negative scale scores, general psychopathology scores, additional criteria scores, and total scores were noted.

We used the HAMA scale (14 items) for estimating anxiety symptomatology (23), and the HAMD scale (24 items) for estimating the severity of the patients' depression (24). Total HAMA and HAMD scores were also noted.

Data analysis

After acquisition, the data were stored in Attribute-Relation File Format (arff) format, native to the Weka data mining package (26), and then analyzed by the Parallel Random Forest (PARF) software (27). This open-source Random Forest (RF) implementation developed at the Ruđer Bošković Institute (Zagreb, Croatia) allows computation to be distributed over a cluster of computers, which becomes relevant when working with large data sets.

We used the RF classifier (28), which has been shown to offer comparable or better classification performance than other state-of-the-art algorithms (18), such as the widely accepted Support Vector Machines (SVM) (29) or C4.5 decision trees (30). Two important features influenced our choice of classifier for this work: 1) its relative insensitivity of the algorithm to parameterization, and 2) the insight into functioning of the RF model by means of an attribute ranking algorithm and construction of class prototypes.

Using the RF classifier, we trained three models for diagnosing PTSD, based on data collected from 1) a structured psychiatric interview; 2) psychiatric symptom scales - CAPS (21), PANSS (22), HAMA (23), HAMD (24); and 3) the combination of the two sources.

The RF classifier (28) is a data mining method based on an ensemble of decision tree models that map observations about an item ("attributes") to conclusions about the item's target value ("class"). An individual decision tree is described by a tree-like structure wherein leaves represent classifications and branches represent conjunctions of features that lead to those classifications. A decision tree can be learned by splitting the source set into subsets based on an attribute value test (31). This process is repeated on each derived subset in a recursive manner. The recursion is completed when splitting is either non-feasible, or a singular classification can be applied to each element of the derived subset. After being "grown" to the maximum, the individual tree is in many implementations – such as the popular C4.5 algorithm (30) – pruned to obtain simpler and more accurate models.

A single RF model consists of a large number of decision trees, each derived from a randomly chosen subset of the data, which then vote to obtain the final classification on the unlabeled instances. The subset is chosen by sampling n out of n instances with replacement. This approach, also called bootstrap sampling, on average picks out $1 - e^{-1} \approx 63.2\%$ of instances for training (32).

The number of attributes available to the algorithm in the construction of each node is also constrained to obtain more robust models (28). Model performance on unseen data are estimated using the out-of-bag procedure, where only the part of the data set not used in construction of a specific tree is used for testing of that tree. This procedure is in concept similar to leave-one-out crossvalidation and should yield comparable results.

The RF can also be used for the estimation of relative attribute importance (Imp). This is achieved by comparing the out-of-bag performance of the model with the performance of the same model run on the data set with a single attribute shuffled. Deterioration in predictive ability signals a relevant attribute. A standardization procedure (28) is used to derive a Z-score (Z-sc) for the extent of such deterioration, allowing its statistical significance to be determined, and corresponding attributes to be ranked. This method of attribute evaluation also allows attributes important only in interaction to be identified. In this work, we supplied the Z-scores of each attribute which may be informative of comparative attribute importance within a single classification experiment, however they cannot be compared between experiments.

All models in this work were built in a two-step fashion: the first iteration was used to measure attribute importance and retain only attributes with *P* values (statistical significance) less than 0.05; those attributes were used in the second iteration to build the final model, estimate its predictive performance on unseen data, construct class prototypes and list relevant attributes. For an attribute to be shown in Tables 1-3, its *P* value has to be below 0.05 also in the second iteration.

The predictive performance of models in this article was normally expressed as accuracy, ie the number of instances for which the class is predicted correctly divided by the total number of instances. Baseline accuracy in our experiments is 50%, meaning a model guessing at random would classify half of all the instances correctly. Other measures of performance used here are sensitivity (proportion of total PTSD cases identified by the model) and specificity (proportion of non-PTSD patients identified by the model).

Using a feature of the PARF software (27), we were able to build specialized models that place more weight on one of the classes (PTSD or non-PTSD), emphasizing either specificity or sensitivity (Table 4). Each shift from the 1:1 ratio does, however, decrease the overall accuracy of the given model, which is expected as 1) the classes are balanced in number and 2) as demonstrated by roughly equal numbers of false positives and false negatives in 1:1 models, classes are equally "difficult to learn." All of the other models described have equally weighted classes (1:1).

Furthermore, we used the option of the PARF software to compute "prototypes" of each class to obtain representative clusters of patients. Clusters are formed using a sample proximity matrix, formed from the statistics obtained by classifying patients using all trees in the forest, and counting the number of times that two samples ended in the same leaf of the tree. Patients that often end up in the same leaves of the trees are close to each other, whereas those that do not are far.

This PARF prototypes feature enables identification of representative sub-groups/clusters of patients of one class, thus providing insight to domain experts and grounds to speculate about possible multiple paths leading to a particular class membership.

Table 1. Relevant attributes in random forest based model 1

Z-score*	Attribute†
165.98	group D of comorbid diagnoses – neurotic, stress-related, and somatoform disorders
102.41	percentage of disability
86.62	duration of military service during the war
79.99	birth order in family
73.66	number of previous hospitalizations
67.84	type of alcohol consumption
65.98	disability status
45.07	physical disability
41.52	period of treatment in outpatient clinic
36.15	duration of current employment status (employment, unemployment, sick leave, retirement)
28.45	age
27.00	duration of life time employment of partner
25.75	year of most intensive psychiatric treatment
24.36	possession of driving license
22.82	duration of current employment status of partner (employment, unemployment, sick leave, retirement)
18.76	relationships in primary family
18.66	number of brothers/sisters
17.90	current employment status (employment, unemployment, sick leave, retirement)
17.71	psychical disability
13.77	personal income
12.18	frequency of alcohol consumption
8.51	group C of comorbid diagnoses - schizophrenia, schizotypal and delusional disorders

*Z-Score – the distance, in units of standard deviation, of a case from the population mean. Here, the population is represented by reductions in number of correctly classified out-of-bag instances for each tree, after a single attribute has been shuffled. The specific case whose distance from the mean this Z-score expresses is zero (no deterioration in accuracy with attribute shuffling, ie a completely irrelevant attribute); therefore, the Z-score is computed by dividing the mean with the standard deviation. Statistical significance is then determined assuming normality.
†Attributes with P value <0.05 are listed, sorted by relevance. The accuracy of this model was 70.59%. Values for standardized value (Z-score) are shown.

Table 2. Relevant attributes in random forest based model 2*

Z-score†	Attribute‡
221.43	CAPS criterion D (hyperarousal symptoms)
190.53	CAPS total score
164.48	PANSS additional criteria score (anger, difficulty in delaying gratification and affective lability)
148.10	group D of comorbid diagnoses – neurotic, stress-related, and somatoform disorders
141.78	CAPS criterion C (avoidance symptoms)
136.19	Hamilton anxiety scale
105.77	CAPS criterion D (hyperarousal symptoms) weighted with validity
104.34	CAPS criterion C (avoidance symptoms) weighted with validity
102.41	CAPS criterion B (re-experiencing symptoms)
95.30	PANSS general psychopathology score
83.64	CAPS total score weighted with validity
55.90	CAPS criterion B (re-experiencing symptoms) weighted with validity
39.87	PANSS total positive score
39.66	Hamilton depression scale
38.48	CAPS criterion A (exposure to a traumatic event)
37.32	PANSS total score
30.33	CAPS inverse validity score
22.29	PANSS total negative score

*Abbreviations: CAPS – Clinician-administered PTSD Scale; PANSS – Positive and Negative Syndrome Scale.
†Z-score – the distance, in units of standard deviation, of a case from the population mean. Here, the population is represented by reductions in number of correctly classified out-of-bag instances for each tree, after a single attribute has been shuffled. The specific case whose distance from the mean this Z-score expresses is zero (no deterioration in accuracy with attribute shuffling, ie a completely irrelevant attribute); therefore, the Z-score is computed by dividing the mean with the standard deviation. Statistical significance is then determined assuming normality.
‡Attributes with P value <0.05 are listed, sorted by relevance. The accuracy of this model was 80.39%. Values for standardized value (Z-score) are shown.

Table 3. Relevant attributes in random forest based model 3*

Z-Score†	Attribute‡
188.59	CAPS criterion D (hyperarousal symptoms)
173.67	CAPS total score
126.02	CAPS criterion C (avoidance symptoms)
118.48	PANSS additional criteria score (anger, difficulty in delaying gratification and affective lability)
113.61	CAPS criterion B (re-experiencing symptoms)
113.17	Hamilton anxiety scale
93.61	CAPS criterion C (avoidance symptoms) weighted with validity
92.82	group D of comorbid diagnoses – neurotic, stress-related, and somatoform disorders
84.96	CAPS criterion D (hyperarousal symptoms) weighted with validity
79.76	CAPS total score weighted with validity
70.33	PANSS general psychopathology score
65.50	CAPS criterion B (re-experiencing symptoms) weighted with validity
57.61	birth order in family
48.56	disability status
41.12	CAPS criterion A (exposure to a traumatic event)
37.59	duration of military service during the war
34.95	percentage of disability
30.13	CAPS inverse validity score
29.36	Hamilton depression scale
27.93	number of brothers/sisters
25.78	duration of current emotional relationship
23.91	PANSS total score
22.48	number of previous hospitalizations
21.31	PANSS total negative score
20.40	number of relatives that patient support
19.20	relationships in primary family
18.37	PANSS total positive score
13.17	physical disability
12.54	duration of current employment status (employment, unemployment, sick leave, retirement)
10.55	age
9.92	period of treatment in outpatient clinic

*Abbreviations: CAPS – Clinician-administered PTSD Scale; PANSS – Positive and Negative Syndrome Scale.
†Z-score – the distance, in units of standard deviation, of a case from the population mean. Here, the population is represented by reductions in number of correctly classified out-of-bag instances for each tree, after a single attribute has been shuffled. The specific case whose distance from the mean this Z-score expresses is zero (no deterioration in accuracy with attribute shuffling, ie a completely irrelevant attribute); therefore, the Z-score is computed by dividing the mean with the standard deviation. Statistical significance is then determined assuming normality.
‡Attributes with P value <0.05 are listed, sorted by relevance. The accuracy of this model was 78.43%. Values for standardized value (Z-score) are shown.

Results

Model 1 – Prediction of diagnosis based on a structured psychiatric interview

The first model was based on the data from a structured psychiatric interview, including the data on main and comorbid diagnosis, family, childhood and personal history, and social and psychiatric data. It included 63 different attributes, 21 of which were shown as relevant (*P* value of the attributes <0.05).

In this model, the most relevant attribute was group D of the comorbid diagnoses, which included neurotic, stress-related, and somatoform disorders (other anxiety disorders, reaction to severe stress and adjustment disorders, and other neurotic disorders) (Table 1). This was followed by attributes that were closely related to the potential for the development of combat PTSD and its consequences (such as percentage of disability, duration of military service during the war, birth order in the family, number of previous hospitalizations) and type of alcohol consumed (Table 1). The out-of-bag crossvalidation accuracy of this model was 70.59%.

Model 2 – Prediction of diagnosis based on psychiatric scales

The attributes included in this model were the main and comorbid diagnosis, along with the results from psychiatric scales – CAPS (A – exposure to a traumatic event, B – re-experiencing symptoms, C – avoidance symptoms, D – hyperarousal symptoms, total scores, validity, weighted scores for B, C, and D criteria, and weighted total scores), HAMA (total scores), HAMD (total scores), PANSS (positive scale scores, negative scale scores, general psychopathology scores, additional criteria scores, total scores). Eighteen of the 22 attributes were relevant (*P* value of the attributes <0.05).

Table 4. Random forest based weighted models*

Data used	Weights (non-PTSD:PTSD)	Specificity (%)†	Sensitivity (%)‡	Accuracy (%)
structured interview	1:1	75	67	0.706
structured interview	2:1	84	35	0.598
structured interview	3:1	98	25	0.618
structured interview	1:2	43	90	0.667
structured interview	1:3	31	96	0.637
psychiatric scales	1:1	80	80	0.804
psychiatric scales	2:1	90	69	0.794
psychiatric scales	3:1	96	59	0.775
psychiatric scales	1:2	65	92	0.784
psychiatric scales	1:3	61	94	0.775
interview + scales	1:1	76	80	0.784
interview + scales	2:1	90	67	0.784
interview + scales	3:1	100	55	0.775
interview + scales	1:2	61	90	0.754
interview + scales	1:3	53	96	0.745

*Models built with different weight for posttraumatic stress disorder (PTSD) or non-PTSD class based on a structured psychiatric interview, psychiatric scales or both.
†Sensitivity – proportion of total PTSD cases identified by the algorithm.
‡Specificity – proportion of non-PTSD cases identified by the algorithm.

The psychiatric scales with comorbid diagnosis had good predictive values. The most relevant attributes were CAPS criterion D (hyperarousal symptoms), along with the total CAPS scores (Table 2). These attributes were followed by PANSS additional criteria (anger, difficulty in delaying gratification, and affective lability) and group D of comorbid diagnoses (neurotic, stress-related, and somatoform disorders – other anxiety disorders, reaction to severe stress and adjustment disorders, other neurotic disorders). The out-of-bag crossvalidation accuracy of this model was 80.39%.

Model 3 – Diagnosis from structured psychiatric interview and psychiatric scales

The third model was trained on the data from both the structured psychiatric interview and psychiatric scales (CAPS, PANSS, HAMA, and HAMD) combined into a single data set. From a total number of 80 attributes, 31 were found to be relevant (*P* value of the attributes <0.05).

After combining all available attributes, CAPS criterion D (hyperarousal symptoms) again surfaced as the most important attribute (Table 3). This was followed by the total CAPS scores, CAPS criterion C (avoidance symptoms), PANSS additional criteria scores (anger, difficulty in delaying gratification, and affective lability), and CAPS criterion B (re-experiencing symptoms). The out-of-bag crossvalidation accuracy of this model was 78.43%.

Additional data mining experiments

Some specialized models were also built by placing more weight on either the PTSD or non-PTSD class (Table 4), sacrificing overall accuracy for possible improvements in specificity (the proportion of non-PTSD cases identified by the model) or sensitivity (the proportion of PTSD cases identified by the model) values.

An example of such a specialized model which combines both the structured psychiatric interview and psychiatric scales, placing three times more weight on the non-PTSD class, achieved 100% specificity (out-of-bag estimate for unseen data), meaning it is highly unlikely to produce false positives.

Finally, we instructed the PARF software to construct prototypes of the PTSD and non-PTSD classes of patients, based on a combination of data from the structured interview and the psychiatric scales, ie the same data used in Model 3, to obtain "representative patients" of each class.

The PTSD class yielded 2 prototype groups, one with 46 and the other with 4 patients. One patient was not assigned to a prototype. The non-PTSD class yielded 3 prototype groups, with 44, 5, and 1 patients described by these groups; the single-patient prototype group was not examined further. Again, one patient was not assigned to a prototype.

Descriptive statistics on the distribution of attribute values among the subgroups are listed in Table 5. Only the attributes shown to be statistically significant (corresponding to attributes in Model 3) are shown, excluding weighted CAPS scores.

For the most relevant attributes, the prototypes of PTSD and non-PTSD classes differed mainly in the results of the scales used, with PTSD patients having higher scores in comparison with non-PTSD patients. Furthermore, non-PTSD patients had group D comorbid diagnoses (neurotic, stress-related, and somatoform disorders) more frequently than PTSD patients.

The two minor prototype groups, with 4 patients for PTSD and 5 patients for the non-PTSD class were distinct from their respective major prototype groups. The scores were lower for the PTSD minor prototype on the psychiatric scales (closer to the non-PTSD group) and neurotic, stress-related, or somatoform disorders were more often diagnosed. The psychiatric scale scores of the minor non-PTSD prototype were higher than for the major one, demonstrating a tendency toward the values observed in the

Table 5. Prototypes of PTSD and non-PTSD class of patients*

		PTSD‡		non-PTSD	
Z-score†	Attributes	Prototype I (n = 46)	Prototype II (n = 4)	Prototype I (n = 44)	Prototype II (n = 5)
188.59	CAPS criterion D score (hyperarousal symptoms)	21 (19-23)	15 (15,15)	13 (0-15)	19 (19,20)
173.67	CAPS total score	64 (57-76)	46 (43-46)	40 (0-49)	60 (53-67)
126.02	CAPS criterion C score (avoidance symptoms)	26 (21-32)	16 (16,16)	12 (0-20)	23 (19-28)
118.48	PANSS additional criteria score (anger, difficulty in delaying gratification and affective lability)	8 (6-10)	6 (5,6)	5 (4-7)	7 (4-7)
113.61	CAPS criterion B score (re-experiencing symptoms)	19 (16-21)	14 (12-14)	10 (0-15)	18 (16-19)
113.17	Hamilton anxiety scale score	21 (17-23)	16 (12-16)	14 (11-16)	19 (15-19)
92.82	group D of comorbid diagnoses – neurotic, stress-related, and somatoform disorders§	majority: 0	majority: 1	majority: 1	majority: 0
70.33	PANSS general psychopathology score	38 (35-44)	34 (33,34)	33 (29-35)	36 (35-37)
57.61	birth order in family	2 (2,2)	2 (2,2)	1 (1,2)	1 (1,2)
48.56	disability status‖	1 (1,2)	2 (1,2)	2 (1-3)	3 (2,3)
41.12	CAPS criterion A (exposure to a traumatic event)¶	1 (1,1)	1 (1,1)	1 (1,2)	1 (1,1)
37.59	duration of military service during the war in years	5 (2-8)	6 (5,6)	2 (0-5)	5 (0-8)
34.95	percentage of disability	20 (0-40)	20 (0-20)	0 (0-0)	0 (0-0)
30.13	CAPS inverse validity score	4 (3,4)	4 (3,4)	4 (4,4)	4 (4,4)
29.36	Hamilton depression scale score	14 (11-21)	10 (8-10)	10 (8-11)	12 (12,13)
27.93	number of brothers/sisters	1 (1,2)	3 (1-3)	1 (1,2)	1 (1,2)
25.78	duration of current emotional relationship in years	11 (0-15)	10 (7-10)	13 (0-19)	20 (14-20)
23.91	PANSS total score	64 (56-78)	55 (52-55)	56 (50-65)	64 (61-65)
22.48	number of previous hospitalizations	2 (0-4)	1 (1,1)	0 (0-1)	0 (0-0)
21.31	PANSS total negative score	9 (8-13)	8 (8,8)	11 (7-14)	14 (11-15)
20.4	number of relatives that patient support	2 (1-3)	1 (1,1)	2 (1,2)	3 (2,3)
19.2	relationships in primary family**	1 (1,1)	1 (1,1)	1 (1,1)	1 (1,2)
18.37	PANSS total positive score	7 (7-16)	7 (7,7)	7 (7,7)	7 (7,7)
13.17	physical disability††	3 (2,3)	3 (3,3)	3 (3,3)	3 (3,3)
12.54	duration of current employment status in years (employment, unemployment, sick leave, retirement)	4 (2-11)	13 (2-13)	3 (1-9)	1 (1-3)
10.55	age in years	38 (35-41)	36 (35,36)	38 (36-48)	43 (37-47)
9.92	period of treatment in outpatient clinic in years	9 (2-11)	7 (4-7)	6 (1-11)	6 (6-11)
9.56	possesion of driving license‡‡	1 (1,1)	1 (1,1)	1 (1,1)	1 (1,1)
5.55	duration of life time employment of partner in years	13 (10-15)	6 (5,6)	13 (10-20)	13 (13-27)
4.99	duration of current employment status of partner in years (employment, unemployment, sick leave, retirement)	12 (9-14)	6 (5,6)	12 (8-20)	12 (12-28)
4.87	current relationship toward primary family§§	1 (1,1)	1 (1,1)	1 (1,2)	1 (1,1)
4.25	legal offenses‖‖	0 (0-1)	1 (0-1)	1 (0-1)	1 (0-1)
2.03	time elapsed from the year of most intensive psychiatric treatment in years	1 (1-7)	1 (1,1)	1 (1,1)	1 (1-6)

*Abbreviations: CAPS – Clinician-administered PTSD Scale; PANSS Positive and Negative Syndrome Scale.
†Z-score – the distance, in units of standard deviation, of a case from the population mean. Here, the population is represented by reductions in number of correctly classified out-of-bag instances for each tree, after a single attribute has been shuffled. The specific case whose distance from the mean this Z-score expresses is zero (no deterioration in accuracy with attribute shuffling, ie a completely irrelevant attribute); therefore, the Z-score is computed by dividing the mean with the standard deviation. Statistical significance is then determined assuming normality.
‡Descriptive statistics of distributions of attribute values for PTSD and non-PTSD class prototype clusters. For numeric attributes (all except group D of diagnoses – neurotic, stress-related, and somatoform disorders) median (5th-95th percentile) is provided. For group D of diagnoses, a categorical attribute, the majority category in the prototype cluster is shown.
§For group D of diagnoses, a categorical attribute, the majority category in the prototype cluster is shown, 0 – not present, 1 – present.
IIDisability status: 1 – disability status present and administratively confirmed; 2 – in the process of realizing administrative conformation of disability status; 3 – disability not present.
¶CAPS criterion A (exposure to a traumatic event): 1 – present; 2 – not present.
**Relationships in primary family: 1 – good; 2 – satisfactory; 3 – poor.
††Physical disability: 1 – permanently; 2 – temporarily; 3 – not present.
‡‡Possesion of driving license: 1 – yes; 2 – No.
§§Current relationship toward primary family: 1 – looking after; 2 – occasionally visiting; 3 – no contact.
IIIILegal offenses: 0 – no offenses; 1 – law offense; 2 – felony; 3 – both 1 and 2.

PTSD group, while the same group of comorbid diagnoses was less frequently diagnosed.

Discussion

We used a feature of the random forest to identify attributes in the structured psychiatric interview and psychiatric scales which showed relevance to the diagnosis of PTSD. To our knowledge, this is one of the first data mining experiments in the field of psychiatry, although significant research has been already done in some medical (10) and non-medical domains (17-19,33).

Our findings showed that the ability of the model to predict PTSD diagnosis based solely on data from the structured psychiatric interview was moderate (out-of-bag crossvalidation accuracy 70.59%). We found high relevance of the group of comorbid diagnoses, which comprised neurotic, stress-related, and somatoform disorders. This was not surprising, considering that many symptoms in this group overlap with the symptoms of PTSD and distinguishing between these groups is often difficult (1). The same group was also shown as highly relevant in the other two models, whereas other groups of comorbid diagnoses were only marginally relevant (such as group C – schizophrenia, schizotypal and delusional disorders) or irrelevant.

Attributes such as the presence and degree of disability, duration of military service during the war, number of previous hospitalizations, and the period of treatment in outpatient clinic represented the cluster of information that related to the conditions and factors leading to the development of combat PTSD and its consequences. Furthermore, duration of military service could also be an indirect measure of duration of exposure to combat (1,34), while the number of previous hospitalizations was also found to be relevant, probably because some institutions did not have other type of treatment programs than institutional treatment.

We found that PTSD patients were significantly more often younger children in the family, indicating that birth order may be a social factor for developing PTSD. Birth order could be related to the development of PTSD in terms of the separation of children from their primary family, family psychopathology, and the reaction of young children to stress events, as their responses are often influenced by the parents' reactions (1).

Finally, alcohol abuse and dependence is often found in PTSD patients (3) as it reduces acute symptomatology, especially the symptoms of anxiety and depression, and could lead to the comorbid diagnosis of alcoholism.

From the second model, based only on psychiatric scales, which was mainly related to the attributes that included current symptoms and conditions, the importance of the CAPS scale surfaced, representing the majority of the relevant attributes in this model. It included both CAPS total scores and CAPS subscales as a measure of main symptoms of PTSD, which is in line with reports of other authors that have also found CAPS scale as a good discriminator for PTSD, using classical statistical methods (35,36). Hyperarousal symptoms were found to be the most relevant CAPS subscale, which are known to be one of the critical symptoms of PTSD, and this finding is in concordance with the hypothesis that noradrenergic system is hyperactive in some patients with PTSD (1). Analyzing other scales, attributes measuring current PTSD symptoms in patients were found important, like the PANSS additional criteria scores (anger, difficulty in delaying gratification and affective lability) and the Hamilton anxiety scale, which was shown to be important in some other studies as well (37), because PTSD is classified as an anxiety disorder.

In the third model, which combined data from both the psychiatric scales and the structured interview, psychiatric scales were clearly identified as more relevant than attributes from the interview. In this model, the diagnoses of neurotic, stress-related, and somatoform disorders were distinguished from PTSD diagnosis. It is important that in both models, CAPS criteria were related in the same order – hyperarousal symptoms were found to be most relevant, followed by avoidance symptoms, re-experiencing symptoms, and finally exposure to a traumatic event.

Taken together, structured interview and psychiatric scales data result in a model of a slightly lower predictive value than the model using psychiatric scales only, with an accuracy of 78.43%. RF has previously been shown to be unhindered by adding weakly predictive attributes (28).

Minor fluctuations in the predictive ability, found in this case, where out-of-bag prediction of only 2 patients changed, would not be considered unusual as they are within the range of inter-experimental variation caused by the stochastic component of RF.

An additional model (data not shown) was also trained to predict PTSD diagnosis from the CAPS scale only, which was shown to be the most useful in PTSD diagnosis. Other psychiatric scales were excluded in this experiment; only a slight deterioration in predictive ability (78.43% out-of-bag classification accuracy) was noted.

Our attempt to weight individual criteria from the CAPS scale with validity values did not yield more important attributes – all non-weighted variables were found to be more relevant than their weighted counterparts. A probable reason for this is that validity of answers is a property of the patients and is essentially unrelated to the CAPS scores itself, and possibly complementary to it in diagnosing PTSD. Therefore, any attempt to combine them into a single value resulted in a loss of information. It is, however, possible that another weighting procedure would better suit our purpose.

Models based on shifting more weight on the PTSD or non-PTSD class could be useful if we need a more specific or sensitive model, with some loss of overall accuracy.

Using prototype models of PTSD and non-PTSD classes in which representatives of each class were described, evident division in results of used scales was found. Minor prototypes that included lower number of patients probably represent "boundary" cases for both the PTSD and non-PTSD classes. Along with that, non-PTSD patients more frequently had some of the comorbid diagnoses that included neurotic, stress-related, and somatoform disorders than the PTSD patients.

All data for our models were acquired on veteran soldiers, and as such all results refer to male patients with combat-related PTSD. It is one of limitations of the study, but the number of women with combat PTSD in Croatia is significantly smaller, ranging from 4 to 6% (5).

In this study, we applied data mining methodology to explore useful patterns that would help us in forming more objective and reliable models for PTSD diagnosis based on collected structured data. Our intent was not to define PTSD itself in detail, but to extract some important attributes that could be of value in the process of decision making.

The classification task in this work could have been addressed by another classifier, such as decision trees or support vector machines. The advantages of the random forest classifier are that it is robust to noise, missing data, and parametrization, and it does not assume that variables are independent, features important in medical and biological research (18). Some authors in biomolecular research showed that it is more critical how the attributes of investigated items were encoded rather than the specific method used (38).

Two possible limitations of the algorithm have been described in the literature: the random forest may have performance inferior to a single pruned decision tree when a very large number of meaningless (noisy) inputs are present (39), an obstacle easily overcome by choice of training parameters. Also, it has been shown that random forests' performance in cases of unbalanced class sizes can be improved (40) by employing class weights or downsampling the majority class during training. None of the two conditions is met by our data set.

In conclusion, using a machine learning approach, we described several attributes important in designing a clinical model for diagnosing PTSD. From experiments conducted on the structured interview data, it seems that, although data about the patient's medical history, and social, economic and marital status may be relevant, indirect data about previous and current symptoms, such as presence and degree of disability, hospitalizations and duration of military service, are of greater importance for the model design. On the other hand, psychiatric scales, which indicate the current condition of a patient, are more relevant in estimating diagnosis, especially the scores on CAPS and its subscales. One more important attribute is the group of comorbid diagnoses comprised of neurotic, stress-related, and somatoform disorders.

The importance of these results is to serve as a proof-of-concept that a data mining approach may be useful in the clinical practice. Also, the trained models and associated attribute rankings could be informative in constructing diagnostic models for PTSD. Future research should include more comprehensive efforts with a larger group of patients, more attributes for description of a given patient and using several different methods of data analysis.

Acknowledgments

Data presented in this article is part of a technological project "Integrative diagnostic model for stress related mental disorders," (TP-03/01, principal investigator Prof. D. Kozarić-Kovačić) supported by the Croatian Ministry of Science, Education, and Sports, which aims to improve diagnostic criteria and develop the diagnostic model for stress related mental disorders that could be used in clinical, pharmacological, forensic, and research procedures. The authors thank Dr Tomislav Šmuc and Dr Dragan Gamberger of the Laboratory for information systems, Ruđer Bošković Institute, for contributing data mining expertise to this work. We also thank Dr Tanja Jovanović for the language revision.

References

1 Sadock BJ, Sadock VA. Kaplan and Sadock's concise textbook of clinical psychiatry. 2nd ed. Baltimore (MD): Williams & Wilkins; 2004.

2 Kozarić-Kovačić D, Marušić A, Ljubin T. Combat-experienced soldiers and tortured prisoners of war differ in the clinical presentation of PTSD. Nord J Psychiatry. 1999;53:11-5.

3 Kozaric-Kovacic D, Kocijan-Hercigonja D. Assessment of post-traumatic stress disorder and comorbidity. Mil Med. 2001;166:677-80. Medline:11515315

4 Kaplan HI, Sadock BJ, Grebb JA. Kaplan and Sadock's Synopsis of Psychiatry. 7th ed. Baltimore (MD): Williams & Wilkins; 1994.

5 Kozaric-Kovacic D, Bajs M, Vidosic S, Matic A, Alegic Karin A, Peraica T. Change of diagnosis of post-traumatic stress disorder related to compensation-seeking. Croat Med J. 2004;45:427-33. Medline:15311415

6 Kozarić-Kovačić D, Borovečki A. Malingering PTSD. In: Corrales TA, editor. Focus on post-traumatic stress disorder research. Hauppauge (NY): Nova Science Publishers, Inc; 2004. p. 185-208.

7 Frueh BC, Hamner MB, Cahill SP, Gold PB, Hamlin KL. Apparent symptom overreporting in combat veterans evaluated for PTSD. Clin Psychol Rev. 2000;20:853-85. Medline:11057375

8 Franklin CL, Repasky SA, Thompson KE, Shelton SA, Uddo M. Differentiating overreporting and extreme distress: MMPI-2 use with compensation-seeking veterans with PTSD. J Pers Assess. 2002;79:274-85. Medline:12425391

9 Sayer NA, Thuras P. The influence of patients' compensation-seeking status on the perceptions of veterans affairs clinicians. Psychiatr Serv. 2002;53:210-2. Medline:11821554

10 Gamberger D, Lavrac N, Krstacic G. Active subgroup mining: a case study in coronary heart disease risk group detection. Artif Intell Med. 2003;28:27-57. Medline:12850312

11 Coulter DM, Bate A, Meyboom RH, Lindquist M, Edwards IR. Antipsychotic drugs and heart muscle disorder in international pharmacovigilance: data mining study. BMJ. 2001;322:1207-9. Medline:11358771

12 Perner P. Intelligent data analysis in medicine-recent advances. Artif Intell Med. 2006;37:1-5. Medline:16338124

13 Frawley WJ, Piatetsky-Shapiro G, Matheus CJ. Knowledge discovery in databases: an overview. AI Magazine. 1992;13:57-70.

14 Hand D, Mannila H, Smyth P. Principles of data mining. Cambridge (MA): MIT Press; 2001.

15 Baker JR, Gamberger D, Mihelcic JR, Sabljić A. Evaluation of artificial intelligence based models for chemical biodegradability prediction. Molecules. 2004;9:989-1004.

16 Hand DJ. Intelligent data analysis: issues and opportunities. Inteligent Data Analysis. 1998;2:67-79.

17 Liu H, Hu ZZ, Torii M, Wu C, Friedman C. Quantitative assessment of dictionary-based protein named entity tagging. J Am Med Inform Assoc. 2006;13:497-507. Medline:16799122

18 Qi Y, Bar-Joseph Z, Klein-Seetharaman J. Evaluation of different biological data and computational classification methods for use in protein interaction prediction. Proteins. 2006;63:490-500. Medline:16450363

19 Goh CS, Lan N, Douglas SM, Wu B, Echols N, Smith A, et al. Mining the structural genomics pipeline: identification of protein properties that affect high-throughput experimental analysis. J Mol Biol. 2004;336:115-30. Medline:14741208

20 American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed. Washington (DC): American Psychiatric Association; 1994.

21 Blake DD, Weathers FW, Nagy LM, Kaloupek DG, Gusman FD, Charney DS, et al. The development of a Clinician-Administered PTSD Scale. J Trauma Stress. 1995;8:75-90. Medline:7712061

22 Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull. 1987;13:261-76. Medline:3616518

23 Hamilton M. The assessment of anxiety states by rating. Br J Med Psychol. 1959;32:50-5. Medline:13638508

24 Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960;23:56-62. Medline:14399272

25 World Health Organization. The ICD-10 classification of mental and behavioral disorders; clinical descriptions and the diagnostic guidelines. Geneva: WHO; 1992.

26 Witten IH, Frank E. Data mining: practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann; 2005.

27 Topić G, Šmuc T. PARF Parallel RF algorithm [computer program]. Zagreb (Croatia): Rudjer Bošković Institute, Center for informatics and computing; 2004. Available from: *http://www.parf.irb.hr*. Accessed: March 26, 2007.

28 Breiman L. Random forests. Machine Learning. 2001;45:5-32.

29 Vapnik VN. The nature of statistical learning theory. Berlin: Springer-Verlag; 1995.

30 Quinlan JR. C4.5: programs for machine learning. San Mateo: Morgan Kaufmann; 1993.

31 Breiman L, Friedman J, Olshen R, Stone C. Classification and regression trees. Belmont (CA): Wadsworth International Group; 1984.

32 Breiman L. Out-of-bag estimation. Berkeley (CA): University of California, Statistics Department; 1998. Available from: *ftp://stat-ftp.berkeley.edu/pub/users/breiman/OOBestimation.ps.Z*. Accessed: March 26, 2007.

33 Hand DJ, Heard NA. Finding groups in gene expression data. J Biomed Biotechnol. 2005;2005:215-25.

34 Buydens-Branchey L, Noumair D, Branchey M. Duration and intensity of combat exposure and posttraumatic stress disorder in Vietnam veterans. J Nerv Ment Dis. 1990;178:582-7. Medline:2394978

35 Hyer L, Summers MN, Boyd S, Litaker M, Boudewyns P. Assessment of older combat veterans with the clinician-administered PTSD scale. J Trauma Stress. 1996;9:587-93. Medline:8827658

36 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;170:558-64. Medline:9330024

37 Kozaric-Kovacic D, Hercigonja DK, Grubisic-Ilic M. Posttraumatic stress disorder and depression in soldiers with combat experiences. Croat Med J. 2001;42:165-70. Medline:11259739

38 Tong W, Hong H, Fang H, Xie Q, Perkins R. Decision forest: combining the predictions of multiple independent decision tree models. J Chem Inf Comput Sci. 2003;43:525-31. Medline:12653517

39 Segal MR. Machine learning benchmarks and random forest regression. San Francisco: University of California, Center for Bioinformatics & Molecular Biostatistics; 2004.

40 Chen C, Liaw A, Breiman L. Using random forest to learn imbalanced data. University of California at Berkeley: Statistics Department; 2004.