Computers in Biology and Medicine 53 (2014) 125–133

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Classification algorithms for the identification of structural injury in TBI using brain electrical activity

Leslie S. Prichep [a,*], Samanwoy Ghosh Dastidar [b], Arnaud Jacquin [b], William Koppes [b], Jonathan Miller [b], Thomas Radman [b], Brian O'Neil [c], Rosanne Naunheim [d], J. Stephen Huff [e,d]

[a] Brain Research Laboratories, Department of Psychiatry, NYU School of Medicine, New York, NY, USA
[b] Algorithm Development Department, BrainScope Co., Inc., Bethesda, MD, USA
[c] Wayne State University, School of Medicine, Department of Emergency Medicine, Detroit, MI, USA
[d] Washington University School of Medicine, Division of Emergency Medicine, St. Louis, MO, USA
[e] Departments of Emergency Medicine and Neurology, University of Virginia, Charlottesville, VA, USA

ARTICLE INFO

Article history:
Received 26 March 2014
Accepted 18 July 2014

Keywords:
TBI
Acute traumatic brain injury
Classifier algorithms
Genetic algorithms
Quantitative brain activity
Electrophysiology of TBI
TBI triage
Structural brain injury
CT+ TBI

ABSTRACT

Background: There is an urgent need for objective criteria adjunctive to standard clinical assessment of acute Traumatic Brain Injury (TBI). Details of the development of a quantitative index to identify structural brain injury based on brain electrical activity will be described.

Methods: Acute closed head injured and normal patients ($n=1470$) were recruited from 16 US Emergency Departments and evaluated using brain electrical activity (EEG) recorded from forehead electrodes. Patients had high GCS (median=15), and most presented with low suspicion of brain injury. Patients were divided into a CT positive (CT+) group and a group with CT negative findings or where CT scans were not ordered according to standard assessment (CT−/CT_NR). Three different classifier methodologies, Ensemble Harmony, Least Absolute Shrinkage and Selection Operator (LASSO), and Genetic Algorithm (GA), were utilized.

Results: Similar performance accuracy was obtained for all three methodologies with an average sensitivity/specificity of 97.5%/59.5%, area under the curves (AUC) of 0.90 and average Negative Predictive Validity (NPV) > 99%. Sensitivity was highest for CT+ cases with potentially life threatening hematomas, where two of three classifiers were 100%.

Conclusion: Similar performance of these classifiers suggests that the optimal separation of the populations was obtained given the overlap of the underlying distributions of features of brain activity. High sensitivity to CT+ injuries (highest in hematomas) and specificity significantly higher than that obtained using ED guidelines for imaging, supports the enhanced clinical utility of this technology and suggests the potential role in the objective, rapid and more optimal triage of TBI patients.

Published by Elsevier Ltd.

1. Introduction

Traumatic brain injury (TBI) accounts for over 1.5 million emergency department (ED) visits annually within the United States and the majority of these visits are for mild injury (mTBI) [1]. The "Center for Disease Control" (CDC) additionally estimates that more than 2 million patients per year with nonfatal traumatic brain injury from sports and recreation activities do not seek medical care [2]. It is critical that medical providers at the scene of the injury and in the ED be able to rapidly and accurately determine if the traumatic brain injury is life or quality of life threatening. Currently, non-contrast cranial computed tomography (CT) scan is the method of choice for evaluating acute brain injury. In the ED departments of US hospitals, patients presenting with mTBI routinely undergo CT scans. This occurs primarily because of the high risk associated with missed intracranial lesions and because current decision rules for the use of CT scanning in mTBI have high sensitivity at the expense of very poor specificity [3].

The emergency medicine, neurosurgical and neuroscience literature indicates that the prevalence of structural brain injury visible on CT ("CT positive", CT+) in adult patients with mTBI and Glasgow Coma Scale score (GCS) [4] of 13–15, evaluated in the ED ranges from 7.5% to 12.1% [5–7]. Several factors related to the status of the patients at the time of evaluation are important

* Correspondence to: Brain Research Laboratories, Old Bellevue Admin. Bldg., 8th Floor, 462 First Avenue, New York, NY 10016, USA. Tel.: +1 212 263 6296; fax: +1 212 263 6457.
E-mail address: Leslie.Prichep@nyumc.org (L.S. Prichep).

http://dx.doi.org/10.1016/j.compbiomed.2014.07.011
0010-4825/Published by Elsevier Ltd.

determinants in the prevalence numbers reported. Furthermore, CT positive scans include a continuum of injuries from those that require little or no clinical intervention to those that are potentially life-threatening, with different clinical paths and actions required for the different levels of injury severity. As defined by Stiell et al. [8] "clinically important" brain lesions are those that require surgical intervention or observation in hospital whereas "clinically unimportant" brain lesions are those that require neither admission nor specialized follow-up.. This underscores the importance of rapid and reliable triage at the point of injury in TBI patients presenting to the ED with low suspicion of brain injury.

Further complicating the situation is the fact that CT scans are assessed by subjective visual inspection, with poor inter-rater reliability, especially for etiology of the injury detected [9,10]. While some methods of scoring the severity of the abnormalities seen in CT have been used, these scales are largely insensitive to the type of injuries seen in patients who present with mild signs and symptoms, (e.g., Marshall Scale) [11]. Thus, there remains a need for more objective quantitative criteria as an adjunct to standard clinical and imaging practice to help optimize sensitivity to the "clinically important" brain injuries.

Changes in brain electrical activity that occur in TBI have been reported in the scientific literature. The use of such measures for the classification and identification of head injured patients has also been studied using 19 lead electroencephalogram (EEG) data, and reported to be sensitive to both structural and functional brain injury [12–14]. This paper will describe the development of a hand-held tool using a classification algorithm based on quantitative features of brain electrical activity for rapid, objective, and reliable assessment of the likelihood of the presence of structural brain injury for the triage of patients who present with mild TBI to the ED. Throughout the period of development independent publications have presented evidence of the potential clinical utility of the classification algorithm in development [15–18].

This paper will describe the development of a quantitative biomarker or index, based on brain electrical activity that is sensitive to the presence of structural brain injury. Embedded in a hand-held device, this system could be used to rapidly, objectively, reliably evaluate head injured patients presenting with low suspicion of brain injury. Specifically, the development and clinical utility of a binary classification algorithm, for classification into one of two distinct categories, will be described. This algorithm would assist in determining whether a patient is likely to have a CT positive brain injury and thereby aid in the determination of the presence of injuries requiring further clinical action.

2. Methods

2.1. Patient population

Data was collected at 16 Emergency Departments (EDs) across the US*, with approval from local Institutional Review Boards (IRBs). Subjects were a convenience sample ($n=1470$; 33% female; 67% male) meeting inclusion/exclusion criteria described below. All subjects signed written informed consent.

2.1.1. Inclusion criteria for TBI patients and controls

These were males and females between the ages of 15 and 90, who suffered a closed head injury and with a GCS of 7 or higher, with or without loss of consciousness (LOC) or traumatic amnesia and with symptoms of TBI. Normal controls were: (1) ED patients presenting without head injuries or problems related to the central nervous system; or subjects who participated in college and high school sports, but who did not sustain head injuries; and (2) possible TBI controls were subjects who sustained head injury but had no altered mental status (AMS), no loss of consciousness (LOC), no amnesia and no significant symptoms related to head injury upon presentation.

2.1.2. Exclusion criteria for TBI patients and normal controls

These were subjects with scalp or skull abnormalities or whose clinical condition, such as head trauma, would not allow placement of the electrodes; intoxication in those obtunded to the point where they could not participate in the study. Patients with advanced dementias, Parkinson's disease, diagnosed chronic drug or alcohol dependence, known seizure disorder, mental retardation, or those currently taking daily prescribed medication for a known diagnosed psychiatric disorder also were excluded.

With regard to factors such as drugs or alcohol, fatigue, pain, and other factors which may be present in head injury cases, the method used in this investigation was to include them in all subject groups (controls and head injured patients), except as defined by exclusion criteria. By doing this, they are eliminated as differentiating factors between groups, and features sensitive to these factors are not selected by the classifier, whereas features independent of such factors that differentiate between groups are candidates for selection.

2.2. Clinical assessments

All study subjects were evaluated with the following symptom based scales or assessment tools: (1) Concussion Symptom Inventory [19] (CSI): a brief screening measure that assesses the presence and severity of 12 common post-concussive symptoms. A Likert-type scale is used to assess symptom severity (range 0–6 per item), with a total score range of 0–72 for the full CSI. Higher scores on the CSI indicate more severe symptoms reported; and (2) Standard Assessment of Concussion (SAC) [20,21]: a brief cognitive screening tool that has been used extensively to assess the cognitive effects of concussion. The SAC includes brief subtests of orientation, immediate memory, concentration, and delayed recall. The total score range of the SAC is 0–30 with lower scores on the SAC indicating poorer cognitive performance. In addition, all measures which were components of the New Orleans Criteria (NOC) for referral of head injured patients for CT scans were collected by trained research assistants. The NOC is a decision guideline for referral for CT in head injured patients which is commonly used in the ED. NOC was computed on each subject for the purpose of estimating the percentage of subjects who would have been referred for CT scan if this decision rule had been used as the sole determinant for referral. The components of the NOC include: headache, vomiting, age above 60 years, drug or alcohol intoxication, persistent anterograde amnesia, anticoagulants, visible trauma above the clavicle, and seizure [6].

Using this information in consultation with emergency medicine and sports medicine physicians, and in conjunction with published guidelines [22,23], the subjects were divided into four clinical categories for the purpose of assessing performance of the classifiers by injury severity. Category 1 subjects were functionally normal controls, Category 2 subjects had mild concussions, Category 3 subjects had moderate concussions, and Category 4 subjects were CT+ (i.e. the CT scan showed evidence of structural injury). Since this study was focused on the binary classification task aimed at discriminating between subjects with structural brain injury positive on CT scan (CT+) from those without structural injury, the first three categories were aggregated into a single category referred to as CT negative or those not referred for a CT (CT−/CT_NR). The age, gender and race distributions of the

sample were determined by the representation of each in the populations served by the participating sites involved.

2.3. CT Scans

The determination of which subjects will receive a CT scan was made by the treating physician according to standard clinical evaluation practice. Results of the CT scans were read by the radiologist (read or over-read by a neuroradiologist) at each clinical site. Based on the CT readings, patients were divided into those with positive findings (CT+) and those with negative findings (CT−). Patients who were not referred for CT (CT_NR) by standard of care were considered together with CT− group as both being not-CT+ in terms of all further analyses. All CT scans were done within 72 hours of injury (with the vast majority done within 24 h).

2.4. EEG data acquisition

Five to ten minutes of eyes closed resting EEG was obtained using BrainScope's prototype handheld EEG acquisition devices. Since the intended use of the TBI algorithm is to aid in the triage of head injured patients in the acute setting (e.g. emergency department), it was important that the electrode sites be the minimum number required to characterize EEG changes in TBI and be easily accessible for rapid set-up. Therefore, EEG was recorded from a limited frontal electrode montage, specifically including the following locations of the expanded International 10–20 Electrode Placement System: Fp1, Fp2, F7, F8, AFz, A1 and A2. Fig. 1 shows the hand-held recording device and the headset used to place electrodes in the required locations. The EEG data was acquired at a sampling rate of 8 kHz. All electrode impedances were below 10 kΩ. Amplifiers had a band pass filter from 0.5 to 70 Hz (3 dB points).

2.5. EEG data processing

Recording sites were re-referenced to linked ears and down-sampled from 8 kHz to 100 Hz prior to any processing of the data. EEG recordings were processed through BrainScope's algorithms for artifact detection [24] in order to identify for removal any biologic and non-biologic contamination, including lateral and horizontal eye movement, EMG muscle activity, high frequency impulse artifacts, extremely low amplitude EEG activity, and atypical electrical activity. In prior studies this artifact detection algorithm showed 87.6% agreement on the artifact-free EEG segments selected by an experienced EEG technologist and the automatic artifact program. Previous experience has demonstrated that sufficient artifact-free data (60–120 s; representing 24–48 contiguous epochs of length 2.56 s) can be obtained from these five to ten-minute recordings in such a population. The epoch length of 2.56 s was selected to correspond to a Fast Fourier Transform (FFT) size of 256 bins used to compute estimates of the power spectra of the subject's EEG.

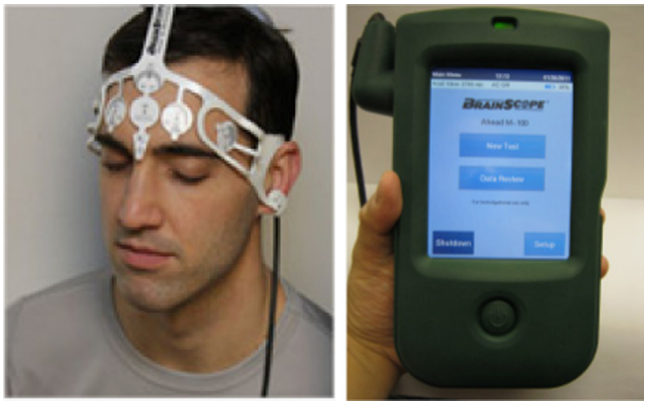

Fig. 1. A patient with the headset for placement of the electrodes in the required standard sites is shown. Hand held EEG data acquisition device is also shown.

2.6. Quantitative EEG (QEEG) feature extraction

The artifact-free EEG data was used to compute quantitative EEG (QEEG) features of absolute and relative power, mean frequency, inter- and intra-hemispheric coherence and symmetry computed for the delta (1.5–3.5 Hz), theta (3.5–7.5 Hz), alpha (7.5–12.5 Hz), beta (12.5–25 Hz) and gamma (30–45 Hz) frequency bands. These measures are described in detail elsewhere [24]. In addition to these traditional QEEG features, several additional features were included: (a) chaotic/fractal measures (fractal dimension and scale-free activity) [25] which evaluate the global complexity of the brain electrical activity at each electrode location across the total spectrum; (b) information theory-based measures (entropy and wavelet entropy), which evaluate the degree of order/disorder of the brain electrical activity at each electrode location; and (c) connectivity measures (phase lag, phase synchrony, and various across-region ratios of spectral power and coherence) which evaluate relationships between and among cortical regions [26,27]

Following neurometric QEEG methodology [28,29] all quantitative features were transformed to obtain an approximately Gaussian distribution and z-transformed relative to age expected normal values. The importance of each of these steps in enhancing the sensitivity and specificity of brain electrical activity has been described in detail elsewhere as are the robust test–retest reliability and independent replications of the normative data base of brain electrical activity [30]. This database of QEEG feature z-scores is referred to in this paper as the "algorithm development database". At this stage, blind to all other data, a final quality check was run to identify non-EEG data contamination based on distribution characteristics of features outside the frequency range of interest that were not identified by the artifact algorithms. Cases with such contamination were eliminated from the database (approximately 3%).

2.7. Feature reduction

Following feature extraction, the algorithm development database contained several thousand features ($M=10{,}308$) and subjects ($n=1470$). The problem of data reduction is common in quantitative electrophysiology as well as in machine learning problems in general. These problems typically involve large datasets where the number of features can be much greater than the number of subjects. In these cases an exhaustive search of such a large feature space for an optimal subset of features from which a classifier is to be constructed becomes computationally prohibitive and statistically limited. Therefore, the first step is to reduce the number of features from several thousand to a few hundred. In this work this problem was approached first using a strategy called "informed feature reduction" (described in detail by Prichep et al. [24]) which was performed to retain only those features that are stable, replicable, physiologically meaningful, and show good separation between the two categories. The reduced feature set following this initial reduction becomes the candidate feature pool for building the classifier.

2.8. Classifier development

The classifier development stage involves searching the reduced dimensionality feature space using advanced machine learning-based search algorithms in order to obtain classifier candidates that can optimally separate the two categories. Fig. 2 shows a flow chart of the process summarized here. Three different classifier builder algorithms were utilized in parallel: Ensemble Harmony (Harmony), Least Absolute Shrinkage and Selection Operator (LASSO), and Genetic Algorithm (GA). In the use of all three methods, the candidates' performance was computed on a total of 35 "splits" of the total dataset. It is noted that the number 35 was arbitrarily chosen and was considered to be well above that demonstrated to be sufficient to get a good estimate of average performance and variability across splits (we have seen anything above or equal to 10 to be adequate). Each "split" is defined as a specific division of the algorithm development database such that a randomly selected 80% of the database forms the "training" portion and the remaining 20% forms the "test" portion. The train and test portions were balanced for category composition. Ten-fold cross-validation performance was used for feature selection using the reduced feature pool as described above. Leave-one-out cross-validation performance was used for selecting the classification threshold. Both of these cross-validation methods are fairly standard and are described in Statistical Classification texts (e.g., [31]). A target cross-validation specificity was used to constrain the solution to acceptable ranges of specificity in accordance with our tolerance for stratification of risk. A decision was made to constrain the specificity instead of the sensitivity because of the larger number of subjects on the "negative side" of the classification task (CT−/CT_NR) which was expected to reduce the chances of inadvertently selecting an operating point based on a spurious peak on the ROC curve. The metrics used for evaluating classifier performance were the area under the curve (AUC) of the ROC as well as sensitivity, specificity, NPV and PPV of the classifiers at their operating point. ROC curves are useful for visualizing classifier performance and the AUC is a simple to compute, global, scalar measure of classifier performance which is commonly used in medical classification problems. In addition, ROC curves have the attractive property of being insensitive to changes in class distributions [32]. At the end of this procedure, an "All-In" classifier building run was conducted on the full algorithm development database (no train and test partitions), with the final objective to utilize the full algorithm development data for training in order to derive a single binary classifier.

While the three classifier builder algorithms all make use of well-established methods for constructing statistical classifiers, they approach the feature subset selection problem in specific ways and impose different constraints. The Ensemble Harmony algorithm and the GA are both evolutionary algorithms. An Evolutionary Algorithm performs a stochastic search (which involves randomness from one iteration to the next) and evaluates a series of candidate solutions, where each new candidate is informed by high-performing predecessors, similar to genetic evolution [33–39]. In addition, the Harmony method uses ensemble classification to

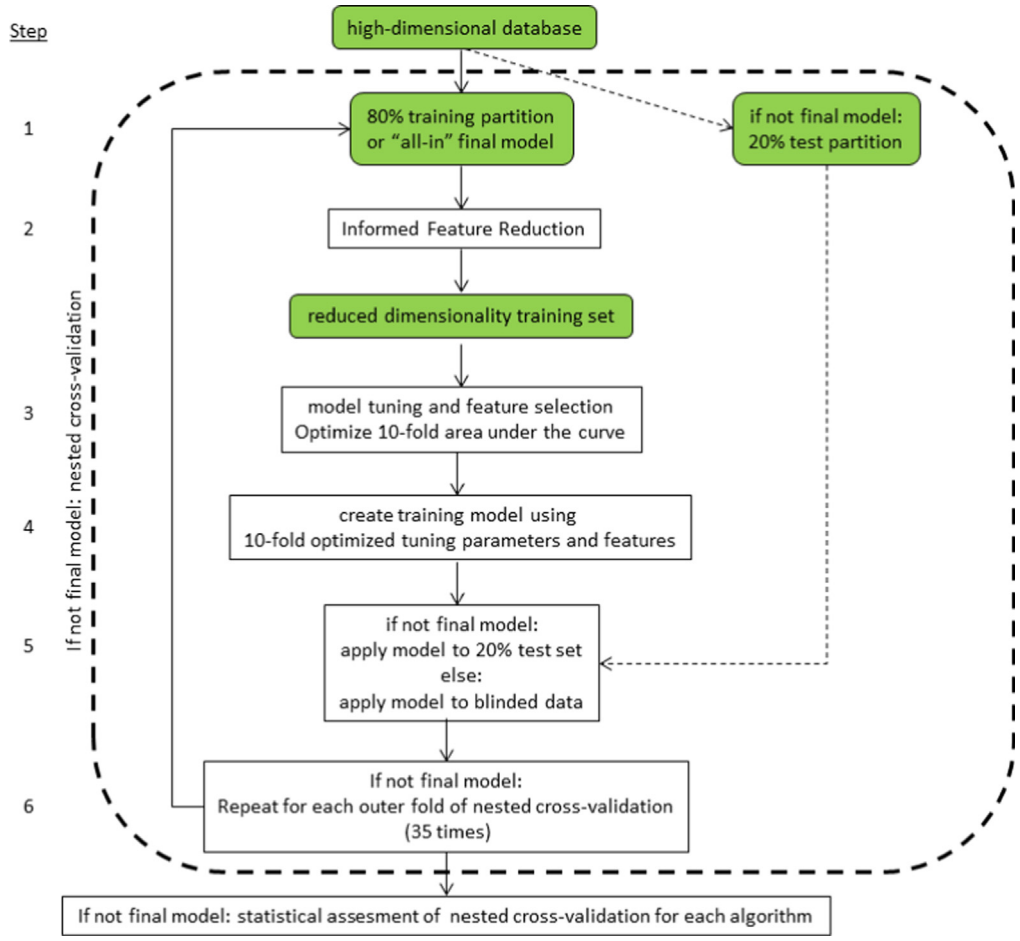

Fig. 2. Flowchart of the general classifier building methodology, including the application of the training process for the nested cross-validation statistical assessment using an 80%/20% train/test partition of the data, repeated 35 times. The final model uses 100% of the data, and is applied to a blinded test group. "If not final model" refers to steps only applied for the nested cross-validation process, but not the final model.

combine a number of evolutionarily derived sub-classifiers (discriminant functions) into a single binary classifier. Combining sub-classifiers in this way serves to average out random anomalies that may be present in any one sub-classifier thereby possibly improving alignment between the combined model and the true underlying signal distributions [40–43]. The *LASSO* method builds a classifier using a regularized logistic regression model with an L1 absolute value ("LASSO") penalty [44]. The L1 penalty increasingly constrains the feature weights, with the optimal constraint level tuned through cross-validation. By adding the LASSO penalty to the starting regression, the original weights are *constrained*, preventing over-fitting. The GA algorithm imposes a penalty constraint on the number of variables in the classifier candidates so as to maintain a high subject-to-variable ratio, thereby reducing the risk of over-fitting. The commonalities and differences in the three methods and their implementation are summarized in Table 1.

3. Results

Subjects ($N=1470$, 973 males and 497 females) were enrolled at 16 Emergency Departments (EDs) across the US, with approval from local Institutional Review Boards (IRBs). The data was divided into two groups based on the presence or absence of structural brain injury: CT−/CT_NR (low suspicion of structural injury based on clinical symptoms) and CT+ (structural injury visible on a CT scan). Table 2 shows the descriptive statistics for the CT−/CT_NR group (1284 subjects) and the CT+ group (186 subjects). The two groups showed significant differences at intake for several demographic and clinical characteristics, with CT+ patients being older, having lower total SAC scores, and more often having history of LOC and AMS related to injury. While GCS for both groups had a median of 15 (normal score), with an interquartile range (IQR) of 0, significant differences were found, possibly reflecting the extremely small variance on the CT−/CT_NR population.

3.1. Classifier performance

The Receiver Operating Characteristic (ROC) curve for the three classifier methods (Harmony, GA and LASSO) generated using the full un-partitioned algorithm development database is shown in Fig. 3. The overall sensitivity/specificity at the selected operating point on the ROC curve (defined by selecting a classification threshold) of the three classifiers were: 97.8/61.4 (Ensemble Harmony), 97.8/59.1 (LASSO), and 96.8/57.9 (GA). The Negative Predictive Value (NPV) and Positive Predictive value (PPV) of the three classifiers were: 99.6/26.8 (Ensemble Harmony), 99.6/25.7 (LASSO), and 99.2/25 (GA). The contingency table for these results is shown in Table 3, and additional performance metrics are shown in Table 4.

As can be seen in Fig. 3, the ROCs for the three classifiers were largely overlapping, demonstrating similar performance on the All-In database, despite the differences in method and selected features. These results suggest that the performance obtained by any one of these methods represented the optimal Bayes decision boundary [45]. This boundary is considered to represent the best possible separation obtainable for a given type of discriminant and a given overlap of the true underlying distributions of the classes, (i.e., CT+ from CT−/CT_NR in this very mild presentation of TBI).

Fig. 4 shows the distribution of discriminant scores for the GA solution as an example of the separation between groups. It can be

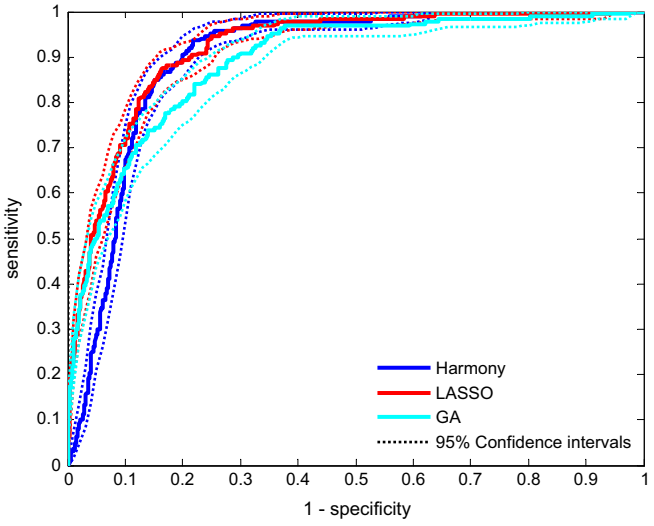

Fig. 3. ROC curves with 95% confidence intervals indicated (dotted lines) for the three classifiers developed using the different classifier building methods (GA, Ensemble Harmony, and LASSO).

Table 1
Commonalities and differences in the three classifier building methods and final classifiers produced by: HARMONY, LASSO, and GA.

	Harmony	LASSO	GA
Classifier building method (feature subset selection)	Stochastic search	Deterministic search	Stochastic search
Discriminant function type	Linear discriminant function (LDF)	Logistic regression	LDF
Variable constraint mechanism	Ensemble of 64 LDFs 20 features each	Weight constraint (L1 norm penalty)	Ceiling for maximum feature number (35)
Objective Function	Area under ROC curve (AUC)	AUC	AUC with penalty for too many features
# of Features in final classifier	152	48	28

Table 2
Descriptive statistics and group comparison *p*-values for subject population enrolled in the study. Where appropriate, median was used and inter-quartile range (IQR, quartile 3 − quartile 1) are reported.

	CT+	CT−/CT_NR	Group comparison	*p*-value
Sample size, *N*	186	1284	Chi square	0.0165
% Female	25.8	35.0	Chi square	0.0128
AGE [Median (IQR, range)]	53.9 (29.0; 18.2–91.7)	32.3 (27.6; 15.1–87.3)	*t*-test	<0.0001
GCS [Median, (IQR; Range)]	15 (0; 7–15)	15 (0; 13–15)	*t*-test	<0.0001
SAC [Median, (IQR; Range)]	23.0 (6; 2–30)	26.0 (5; 7–30)	*t*-test	<0.0001
LOC (% positive)	72.0	39.2	Chi square	<0.0001
AMS (% positive)	51.1	31.1	Chi square	<0.0001

seen that the discriminant scores for the CT−/CT_NR group (shown in green) lies to the right of the both the CT+ groups, shown in blue (CT+ but not hematomas) and red (CT+ hematomas). The discriminant score distribution of the "most severe group" (group with brain injury which is potentially most life-threatening), seen in red, contains the most abnormal scores. These findings support the fact that the GA classifier is sensitive to the continuum of severity present within the CT+ population. It is noted that similar distributions were obtained for the other two classifiers.

In order to assess the impact of severity of injury on the classification performance, we computed the percentage of subjects classified as CT+ for five functional/structural injury categories of subjects: these were the three categories previously defined in Section 2.2 [categories 1 (controls), 2 (mild concussion), 3 (moderate concussion)], along with the following two subgroups of category 4 subjects: category 4 "clinically unimportant" structural injury, and category 4 "clinically important" traumatic hematomas with measureable blood. The percentage of these subjects classified as CT+ is summarized in Table 5. Note that the number of traumatic hematomas in category 4/CT+ subjects is quite small ($n=12$ and therefore, one miss-classification appears to be a large performance drop even though it is insignificant. For all three classifiers, it can be seen that there was a clear relationship between the degree of functional impairment and classifier performance. For example, in the case of the harmony classifier, 31.0% of category 1, 40.3% of category 2, 46.9% of category 3 subjects, and 97.8% of category 4 subjects were classified as CT+. Furthermore, for the clinically most important group (hematomas) performance was 100% for both Harmony and LASSO. This is further illustrated in Fig. 4 which shows the distribution of discriminant scores for the GA algorithm, color coded for clinical categories 1–3 (CT−/CT_NR), category 4 (CT+ which are not severe) and category 4 hematomas (CT+ severe). Greater overlap can be seen between the CT−/CT_NR and "clinically unimportant" non-severe CT+ cases, with greatest separation for the severe CT+ cases.

The three classifier models selected different QEEG features and a different number of features, supporting the premise that there is no one unique solution to the optimal separation of the two groups (CT+ and CT−/CT_NR). However, the features contributing most to classification for all three on the artifact-free EEG segments selected by included features from the same feature domains (measure sets). The selected features for all classifiers contained measures which reflect changes in power and frequency distributions, largely reflected in shifts in mean frequency of the total spectrum, as well as individual wide bands, especially alpha. The multivariate expression of these features reflecting the side of maximal abnormality, with no meaning attributed to which hemisphere the maximum occurred on, played a pivotal role. Also important were features that measure disturbances in connectivity between regions (including coherence and phase synchrony), and ratios of these quantities, in order to capture temporal and spatial relationships in brain activity among different regions and frequency bands.

Table 6 is the contingency table for the NOC applied to this population. It can be seen that the specificity of the NOC was extremely low for the CT−/CT_NR patients. Comparing these numbers to the specificity of the TBI algorithm shows the improvement of the algorithm over the standard of care using a decision rule like the NOC.

Table 3
Contingency tables for the three binary classifiers.

Predicted		True		
		CT−/CT_NR	CT+	Severe CT+
Harmony	**CT−/CT_NR**	788	4	0
	CT+	496	182	12
GA	**CT−/CT_NR**	744	6	1
	CT+	540	180	11
LASSO	**CT−/CT_NR**	759	4	0
	CT+	525	182	12

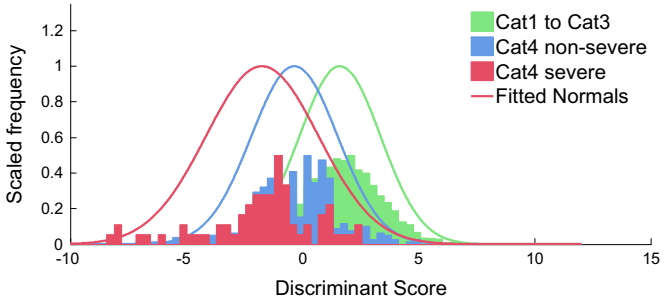

Fig. 4. Histograms of the GA discriminant scores with fitted Gaussians, as a function of clinical category, shown in green for categories 1, 2 and 3 (CT−/CT_NR), blue for CT+ cases that were not hematomas and in red for the hematoma subgroups ("clinically important") of the CT+ patients. The histograms are each independently scaled to have peak value of 0.5, and the fitted curves are scaled to have peak value of 1.0. High scores correspond to normal patients and low scores correspond to abnormality. The threshold for classifying Cat4 is set at 0.637. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Comparison of performance metrics for all three classifiers. Metrics of performance reported in the Table include: Sensitivity/Specificity (with 95% confidence intervals (CI)), Area Under the ROC Curve (AUC, with 95% CI), Negative Predictive Value (NPV) and Positive Predictive Value (PPV) and Sensitivity on the subgroup of "Severe CT+" ($N=12$). For calculation of NPV and PPV the 10% prevalence rates are used.

Comparative performance	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	AUC (95% CI)	NPV	PPV	CT+ Severe sensitivity (%)
Harmony	97.85 (94.23–99.31)	61.37 (58.64–64.03)	0.90 (0.89–0.91)	99.6	26.8	100.00
LASSO	97.85 (94.23–99.31)	59.11 (56.36–61.81)	0.92 (0.90–0.93)	99.6	25.7	100.00
GA	96.77 (92.79–98.68)	57.94 (55.18–60.65)	0.89 (0.87–0.91)	99.2	25.0	91.7

Table 5
Impact of severity of functional impairment on classifier performance; Percentage of subjects classified as CT+ for five categories of subjects with increasing degree of functional impairment (starting from Cat 1 which are the normal control subjects).

	Percentage of subjects classified as CT+				
	Cat. 1	Cat. 2	Cat. 3	Cat. 4 Non-severe	Cat. 4 Severe
N	497	414	373	174	12
Harmony	30.99	40.34	46.92	97.73	100.00
LASSO	33.40	42.27	49.33	97.73	100.00
GA	35.41	43.48	49.33	97.16	91.67

Table 6
Contingency table applying the NOC to all study subjects. It is noted that all subjects did not have all the necessary items to complete the NOC, the table shows those with all items recorded (~75%).

	NOC+	NOC−	Sensitivity	Specificity
CT+	181	5	97.30%	na
CT−/CT_NR	881	72	na	7.56%

4. Discussion

Three different classifier building algorithms were used to explore the optimal separation of CT+ and CT−/CT_NR head injured patients using subsets of quantitative features of brain electrical activity. The three classifiers used different approaches to data reduction and feature selection and imposed different constraints. All three classifiers reported high sensitivity (average 97.5%) and specificity (average 59.5%) for traumatic structural brain injury obtained using the classification approaches described in this work. The NOC in the study sample (for those with all items available) had a sensitivity equivalent to the three classifiers, but on the artifact-free EEG segments selected by an extremely low specificity of less than 8%, while the classifiers had a specificity of approximately 60%, greatly exceeding that of the currently used symptom based decision rules. It should be noted that due to the high risk associated with false negative results, the algorithm performance was specifically targeted to maximize sensitivity and NPV at the expense of specificity and PPV, but while still achieving specificity and PPV results that outperform currently available standard of care assessments. These results suggest that the BrainScope classifier used as an adjunct to standard clinical evaluation has the potential to greatly reduce unnecessary exposure to radiation associated with CT scans. Extremely high sensitivity to hematomas was found without limitations to hematoma volume which constrain current existing tools (e.g., those based on near infra-red spectroscopy).

The increased risk for brain injury in the older population when a head injury is sustained is consistent with our finding of a higher age in the CT+ population. Since all variables are age regressed relative to age expected normal values, the effect of age per se is removed as a factor from the features derived from the brain electrical activity data. In addition, no significant correlations were found between age and classification accuracy. The increased incidence of LOC and AMS in the CT+ population and the lower mean SAC score were not unexpected, as the severity of the injury would support such differences. While significant differences were reported for GCS, it is more important to note that both groups have a median of 15 (normal score) and an interquartile range (IQR) of 0, giving support to the fact that both groups were characterized by a GCS in the normal range, tightly around the highest score of 15. The fact that there are a small number of CT+ patients with lower GSC scores, and fewer outliers in the CT-/CT_NR population, suggests that the variance differences explain the significant, although not clinically meaningful, finding.

The observed similar performance of the three classifiers suggests that the accuracy obtained by any one of these methods represents the optimal Bayes decision boundary [45]. This boundary is considered to represent the best possible separation obtainable, for a given type of discriminant and a given overlap of the true underlying distributions of the classes, (i.e., CT+ from CT−/CT_NR in this very mild presentation of TBI). Related to the overlap of these distributions is the subjective nature of visual evaluation and poor inter-rater reliability of the "gold standard" CT scan particularly in milder forms of TBI. It is also important to take into account the nature and severity of the structural brain damage indicated by a positive CT scan. CT+ scans include a continuum of injury from those that require little or no clinical intervention to those that are clinically important and potentially life-threatening, with associated different clinical responses and actions required. The importance of additional quantitative evaluations of the CT scans was demonstrated by the extremely high sensitivity obtained by the algorithms when a measure of blood volume in hematomas was used to define a "clinically important" subgroup of the CT+ TBI population, where two of the three classifiers had sensitivity of 100% and the third has a sensitivity of 92%. Furthermore, the high sensitivity of the algorithms to the full spectrum of CT+ injuries suggests possible benefit even when neurosurgery is not involved, and might aid in appropriate triage and interventions such as restriction from sports activities, participation in vigorous military duties, or additional clinical observation. This technology may be especially relevant in situations where CT scans are not readily available and transport for treatment is needed such as in sports or military settings.

The pathophysiology of TBI is complex and related to many different aspects of brain function, including neurometabolic, neurophysiological and structural changes [46]. Quantitative features of brain electrical activity (QEEG) used in the BrainScope technology have also been shown to be sensitive to these changes in brain activity [12–14] without the limitations of other neuroimaging tools (e.g., availability at point of care, radiation exposure, cost-effectiveness). For example, the hypometabolism reported in PET imaging in TBI is consistent with slowing of the EEG spectra seen in this population. Furthermore, changes in connectivity reported in TBI using Diffusion Tensor Imaging (DTI) are consistent with the phase synchrony abnormalities reported using QEEG [47]. It is also important to point out that there is not one unique solution to the classification problem addressed in this work. Rather, there is a multivariate profile of features which demonstrate separation. Since there is a high correlation among EEG features, especially in the restricted number of brain regions sampled with this device, the important factor is whether the three methods drew features from similar measures sets which describe the underlying nature of the abnormalities detected. When evaluated by measure set it is noted that using the features selected by the GA (the smallest number of features used), the overlap with Harmony is 92% and with LASSO is 75%. Thus, while the three classifier models described selected different features and different numbers of features, the feature sets in all three were representative of similar alterations in brain function and well reflected the underlying pathophysiology hypothesized in the scientific literature for structural and functional brain injuries sustained in TBI.

A limitation of the current study is that we did not acquire follow-up data from patients after ED discharge. Future studies would be desired to investigate expanded applications of this technology for following recovery and progression of TBI to further optimize treatment and minimize unnecessary follow-up imaging. Another limitation is the small number of hematoma patients in the CT+ population, a replication in a larger population would be desirable. The need for independent validation of the predicted performance of these classifiers is necessary and is currently underway.

Clinical results obtained using the classifier algorithms constructed using the methodology described in this paper demonstrate the potential added value of using such technology as an adjunct to current clinical practice to assist in more objective, timely and optimal triage of mTBI patients who have structural brain injury. In addition, the specificity of the classifiers has been shown to be better than that for existing standard guidelines for determination of referral of head injured patients for CT scan, thus potentially impacting on unnecessary CT scans.

5. Summary

There is an urgent need for objective criteria as an adjunct to standard clinical assessment for the identification of acute Traumatic Brain Injury (TBI). Details of the development of a quantitative index based on brain electrical activity that is highly sensitive to the presence of structural brain injury is described. Closed head injured and normal patients ($n=1470$) were recruited from 16 U.S. Emergency Departments (ED) and evaluated within 72 h of injury, using brain electrical activity (EEG) recorded from a limited montage of electrodes on the frontal and frontotemporal regions of the forehead and given a short battery of clinical/neurocognitive tests. At time of the evaluation patients had high GCS (median=15), and most presented with low suspicion of brain injury. Following radiological studies patients were divided into two groups, a CT positive (CT+) group and a group with CT negative findings or where CT scans were not considered to be needed according to current standard of care (CT−/CT_NR). Three different classifier builder methods were utilized, including: Ensemble Harmony, Least Absolute Shrinkage and Selection Operator (LASSO), and Genetic Algorithm (GA). Input to the algorithms was a selected subset of linear and non-linear features describing brain electrical activity in terms of power, connectivity and complexity. All features were *z*-transformed relative to age expected normal values and transformed to obtain Gaussianity. Similar performance was obtained for all three classifier methodologies with an average sensitivity/specificity of 97.5%/59.5%, with area under the curves (AUC) of 0.90 and average Negative Predictive Validity (NPV) greater than 99%. Sensitivity was higher for CT+ cases with potentially life threatening hematomas, where two of three classifiers were 100%. Applying the New Orleans's Criteria (NOC), a guideline for CT referral in TBI used often in the ED, to the study population, specificity was 7.6%. The observed similar performance of these classifiers suggests that the accuracy obtained represents the optimal separation obtainable given the overlap of the underlying distributions of brain activity measures within the CT+ and CT-/CT_NR populations. High sensitivity to CT+ injuries (highest in clinically important hematomas) and specificity significantly higher than that obtained by widely used standard guidelines for imaging in the ED, supports the enhanced clinical utility of this methodology and suggests the potential role of such a technology in the objective, rapid and more optimal assessment and triage of TBI patients.

Conflict of interest

This research was supported by a clinical research grant from BrainScope, Co., Inc., to the clinical sites participating in data acquisition. Three of the co-authors, Drs. Naunheim, O'Neil and Huff, are principal investigators at their respective sites. Dr. Prichep serves as consultant to BrainScope, Inc., and holds financial interest in BrainScope, Co., Inc. through patented technology from NYU School of Medicine. Drs. Ghosh-Dastidar, Jacquin, and Radman, and Mr. Koppes and Mr. Miller are BrainScope employees and members of the BrainScope Algorithm Development team. Strict adherence to ethical concerns was followed and all subjects signed written informed consents for participation in the study.

Acknowledgements

The authors acknowledge the contributions of those who made this research possible, including the primary investigators and research technicians at all the clinical sites and the patients who volunteered. The authors also acknowledge the contributions of Greg Mahowald and Julie Filipenko in support of the project. This research was supported by BrainScope Co., Inc., Bethesda Maryland who granted expenses related to data acquisition.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at: http://dx.doi.org/10.1016/j.compbiomed.2014.07.011.

References

[1] M. Faul, L. Xu, M.M. Wald, V.G. Coronado, Traumatic Brain Injury in the United States: Emergency Department Visits, Hospitalizations, and Deaths 2002–2006, Atlanta, 2010.
[2] J.A. Langlois, W. Rutland-Brown, M.M. Wald, The epidemiology and impact of traumatic brain injury: a brief overview, J. Head Trauma Rehabil. 21 (5) (2006).
[3] A.S. Jagoda, J.J. Bazarian, J.J. Bruns Jr, S.V. Cantrill, A.D. Gean, P.K. Howard, J. Ghajar, S. Riggio, D.W. Wright, R.L. Wears, A. Bakshy, P. Burgess, M.M. Wald, R.R. Whitson, Clinical Policy: neuroimaging and decisionmaking in adult mild traumatic brain injury in the acute setting, Ann. Emerg. Med. 52 (6) (2008) 714–748.
[4] G. Teasdale, B. Jennett, Assessment of coma and impaired consciousness, Lancet 2 (1974) 81–83.
[5] I.G. Stiell, C.M. Clement, B.H. Rowe, M.J. Schull, R. Brison, D. Cass, M. A. Eisenhauer, R.D. McKnight, G. Bandiera, B. Holroyd, J.S. Lee, J. Dreyer, J. R. Worthington, M. Reardon, G. Greenberg, H. Lesiuk, I. MacPhail, G.A. Wells, Comparison of the Canadian CT Head Rule and the New Orleans Criteria in patients with minor head injury, J. Am. Med. Assoc. 294 (12) (2005) 1511–1518.
[6] M. Smits, D.W.J. Dippel, G.G. de Haan, H.M. Dekker, P.E. Vos, D.R. Kool, P. J. Nederkoorn, P.A.M. Hofman, A. Twijnstra, H.L.J. Tanghe, M.G.M. Hunink, External validation of the Canadian CT Head Rule and the New Orleans Criteria for CT Scanning in patients with minor head injury, J. Am. Med. Assoc. 294 (12) (2005) 1519–1525.
[7] J. Ibanez, F. Arikan, S. Pedraza, E. Sanchez, M.A. Poca, D. Rodriguez, E. Rubio, Reliability of clinical guidelines in the detection of patients at risk following mild head injury: results of a prospective study, J. Neurosurg. 100 (2004) 825–834.
[8] I.G. Stiell, H. Lesink, K. Vandemheen, C. Clement, M. Reardon, M.A. Eisenhauer, Obtaining consensus for the definition of 'Clinically Important' brain injury in the CCC Study, Acad. Emerg. Med. 7 (2000) 572.
[9] D.K. Molina, J.J. Nichols, V.J. DiMaio, The sensitivity of computed tomography (CT) scans in detecting trauma: are CT scans reliable enough for courtroom testimony? J. Trauma 63 (2007) 625–629.
[10] J.P. Laalo, T.J. Kurki, P.H. Sonninen, O.S. Tenovuo, Reliability of diagnosis of traumatic brain injury by computed tomography in the acute phase, J. Neurotrauma 26 (2009) 2169–2178.
[11] L.F. Marshall, S. Bowers-Marshall, M.R. Klauber, M. van Berkum Clark, H. M. Eisenberg, J.A. Jane, T.G. Luerssen, A. Marmarou, M.A. Foulkes, A new classification of head injury based on computerized tomography, J. Neurosurg. 75 (Suppl.) (1991) S14–S30.
[12] M.T. Tebano, M. Cameroni, G. Gallozzi, A. Loizzo, G. Palazzino, G. Pezzini, G. F. Ricci, EEG spectral analysis after minor head injury in man, Electroencephalogr. Clin. Neurophysiol. 70 (2) (1988) 185–189.
[13] R.W. Thatcher, C. Biver, R. Mc Alaster, M. Camacho, A. Salazar, Biophysical linkage betweeen MRI and EEG amplitude in closed head I, Neuroimage 7 (1998) 352–367.
[14] R.W. Thatcher, D.M. North, R.T. Curtin, R.A. Walker, C.J. Biver, J.F. Gomez, A. M. Salazar, An EEG severity index of traumatic brain injury, J. Neuropsychiatry Clin. Neurosci. 13 (1) (2001) 77–81.
[15] R.S. Naunheim, M. Treaster, J. English, T. Casner, R. Chabot, Use of brain electrical activity to quantify traumatic brain injury in the Emergency Department, Brain Inj. 24 (11) (2010) 1324–1329.
[16] R.S. Naunheim, M. Treaster, J. English, T. Casner, Automated electroencephalogram identifies abnormalities in the ED, Am. J. Emerg. Med. 29 (8) (2011) 845–848.
[17] B. O'Neil, R.S. Naunheim, L.S. Prichep, R.J. Chabot, Can quantitative brain electrical activity aid in the initial screening of mild traumatic brain injured patients, West. J. Emerg. Med. 13 (5) (2012) 394–400.
[18] D.F. Hanley, R.J. Chabot, W.A. Mould, T. Morgan, R.S. Naunheim, K. Sheth, W. Chiang, L.S. Prichep, Use of brain electrical activity for the identification of hematomas in mild traumatic brain injury (epub ahead of print), J. Neurotrauma (2013), http://dx.doi.org/10.1089/neu.2013.3062.
[19] C. Randolph, S. Millis, W.B. Barr, M. McCrea, K.M. Guskiewicz, T.A. Hammeke, J. P. Elly, Concussion Symptom Inventory: an empirically derived scale for monitoring resolution of symptoms following sport-related concussion, Arch. Clin. Neuropsychol. 24 (3) (2009) 219–229.
[20] M. McCrea, W.B. Barr, K. Guskiewicz, C. Randolph, S.W. Marshall, R. Cantu, J. A. Onate, J.P. Kelly, Standard regression-based methods for measuring

recovery after sport-related concussion, J. Int. Neuropsychol. Soc. 11 (1) (2005) 58–69.
[21] M. McCrea, K.M. Guskiewicz, S.W. Marshall, W. Barr, C. Randolph, R. Cantu, J. A. Onate, J. Yang, J.P. Kelly, Acute effects and recovery time following concussion in collegiate football players: the NCAA Concussion Study, JAMA 290 (2003) 2556–2563.
[22] R.C. Cantu, Posttraumatic retrograde and anterograde amnesia: pathophysiology and implications in grading and safe return to play, J. Athl. Train. 36 (3) (2001) 244–248.
[23] D. Erlanger, T. Kaushik, R. Cantu, J.T. Barth, D.K. Broshek, J.R. Freeman, F. M. Webbe, Symptom-based assessment of the severity of a concussion, J. Neurosurg. 98 (3) (2003) 477–484.
[24] L.S. Prichep, A. Jacquin, J. Filipenko, S. Ghosh Dastidar, S. Zabele, A. Vodencarevic, N.S. Rothman, Classification of traumatic brain injury severity using informed data reduction in a series of binary classification algorithms, IEEE Trans. Neural Syst. Rehabil. Eng. 20 (6) (2012) 806–822.
[25] B.J. He, J.M. Zempel, A.Z. Snyder, M.E. Raichle, The temporal structures and functional significance of scale-free brain activity, Neuron 66 (3) (2010) 353–369.
[26] N.V. Thakor, S. Tong, Advances in quantitative electroencephalogram analysis methods, Annu. Rev. Biomed. Eng. 6 (1) (2004) 453–495.
[27] V. Sakkalis, Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG, Comput. Biol. Med. 41 (12) (2011) 1110–1117.
[28] E.R. John, L.S. Prichep, J. Friedman, P. Easton, Neurometrics: computer-assisted differential diagnosis of brain dysfunctions, Science 293 (1988) 162–169.
[29] E.R. John, L.S. Prichep, P. Easton, Normative data banks and neurometrics: basic concepts, methods and results of norm construction, in: A.S. Gevins, A. Remond (Eds.), Handbook of Electroencephalography and Clinical Neurophysiology, vol. I, Elsevier, Amsterdam, 1987, pp. 449–495.
[30] L.S. Prichep, Use of normative databases and statistical methods in demonstrating clinical utility of QEEG: importance and cautions, Clin. EEG 36 (2) (2005) 82–87.
[31] O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, second edition, John Wiley and Sons, Inc., New York, 2001.
[32] T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (8) (2006) 861–874.
[33] M. Tomassini, Parallel and distributed evolutionary algorithms: a review, in: K. M. M.M. J.P. P. Neittaanmki (Ed.), Evolutionary Algorithms in Engineering and Computer Science J, Wiley and Sons, Chichester, 1999.
[34] G. Syswerda, J. Palmucci, The application of genetic algorithms to resource scheduling, in: Proceedings of the Fourth International Conference on Genetic Algorithms, University of California, San Diego, 13–16 July 1991, pp. 502–508.
[35] M. Mitchell, An Introduction to Genetic Algorithms (Complex Adaptive Systems), First MIT Press, Cambridge, 1996.
[36] W. Siedlecki, J. Sklansky, A note on genetic algorithms for large-scale feature selection, Pattern Recognit. Lett. 10 (5) (1989) 335–347.
[37] J.H. Yang, V. Honavar, Feature subset selection using a genetic algorithm, IEEE Intell. Syst. 13 (2) (1998) 44–49.
[38] M.L. Raymer, E.D. Punch, E.D. Goodman, L.A. Kuhn, A.K. Jain, Dimensionality reduction using genetic algorithms, IEEE Trans. Evolut. Comput. 4 (2) (2000) 164–171.
[39] I.S. Oh, J.S. Lee, B.R. Moon, Hybrid genetic algorithms for feature selection, IEEE Trans. Pattern Anal. Mach. Intell. 26 (11) (2004) 1424–1437.
[40] J.A. Baranauskas, M.C. Monard, Combining symbolic classifiers from multiple inducers, Knowl.-Based Syst. 16 (3) (2003) 129–136.
[41] L. Rokach, Ensemble-based classifiers, Artif. Intell. Rev. 33 (1–2) (2010) 1–39.
[42] R. Polikar, Ensemble based systems in decision making, Circuits Syst. Mag. IEEE 6 (3) (2006) 21–45.
[43] D. Optiz, R. Maclin, Popular ensemble methods: an empirical study, J. Artif. Intell. Res. 11 (1999) 169–198.
[44] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Statist. Soc. B 58 (1) (2013) 267–288.
[45] K. Tumer, J. Ghosh, Estimating the Bayes error rate through classifier combining, in: Proceedings of the 13th International Conference on Pattern Recognition, 1996, pp. 25–29.
[46] C.C. Giza, D.A. Hovda, The neurometabolic cascade of concussion, J. Athl. Train. 36 (2001) 228–235.
[47] A.R. Sponheim, K.A. McGuire, S.S. Kang, N.D. Davenport, S. Aviyente, E. M. Bernat, K.L. Lim, Evidence of disrupted functional connectivity in the brain after combat-related blast injury, Neuroimage 54 (2011) s21–s29.