# Classification of Traumatic Brain Injury Severity Using Informed Data Reduction in a Series of Binary Classifier Algorithms

Leslie S. Prichep, Arnaud Jacquin, *Member, IEEE*, Julie Filipenko, Samanwoy Ghosh Dastidar, Stephen Zabele, Asmir Vodenčarević, *Member, IEEE*, and Neil S. Rothman

***Abstract*—Assessment of medical disorders is often aided by objective diagnostic tests which can lead to early intervention and appropriate treatment. In the case of brain dysfunction caused by head injury, there is an urgent need for quantitative evaluation methods to aid in acute triage of those subjects who have sustained traumatic brain injury (TBI). Current clinical tools to detect mild TBI (mTBI/concussion) are limited to subjective reports of symptoms and short neurocognitive batteries, offering little objective evidence for clinical decisions; or computed tomography (CT) scans, with radiation-risk, that are most often negative in mTBI. This paper describes a novel methodology for the development of algorithms to provide multi-class classification in a substantial population of brain injured subjects, across a broad age range and representative subpopulations. The method is based on age-regressed quantitative features (linear and nonlinear) extracted from brain electrical activity recorded from a limited montage of scalp electrodes. These features are used as input to a unique "informed data reduction" method, maximizing confidence of prospective validation and minimizing over-fitting. A training set for supervised learning was used, including: "normal control," "concussed," and "structural injury/CT positive (CT+)." The classifier function separating CT+ from the other groups demonstrated a sensitivity of 96% and specificity of 78%; the classifier separating "normal controls" from the other groups demonstrated a sensitivity of 81% and specificity of 74%, suggesting high utility of such classifiers in acute clinical settings. The use of a sequence of classifiers where the desired risk can be stratified further supports clinical utility.**

***Index Terms*—Genetic algorithms (GAs), informed data reduction, multiclass classification, quantitative electroencephalography (QEEG), traumatic brain injury (TBI).**

## I. Introduction

ASSESSMENT of medical disorders is often aided by objective diagnostic tests used for early intervention and appropriate treatment. Whether integrated into a testing device or as a comparison with normal ranges for interpretation of test results, classification of subjects into two or more categories is an integral part of the diagnostic process. Many types of classification approaches have been used for medical diagnostics, including discriminant analysis, logistic regression, cluster analysis, neural networks, and tree-based algorithms [1], [2]. The method selected for analyzing and developing classification algorithms for diagnostic data must take into account the size of the dataset (sample size and number of features) and the construction and dimensionality of the feature space [3]. In the field of neurological and psychiatric disorders, there is an extensive literature demonstrating the clinical utility of linear discriminant functions or other types of classifier functions, using features extracted from quantitative analysis of the electroencephalography (QEEG), as adjuncts to the diagnostic process, early detection, and treatment optimization [4]–[7].

This paper focuses on the derivation of a method for the objective evaluation and classification of traumatic brain injury (TBI) subjects using quantitative features of brain electrical activity, with particular emphasis on the translation of these methods into clinically useful tools for the acute care setting. There are more than 1.7 million emergency department (ED) visits annually for TBI [8] an estimated 75% of which are "mild," and an estimated 3.8 million sports related concussive injuries incurred annually in the United States [9]. Of critical importance to the emergency medicine clinician, whether in a civilian or military environment, is the ability to identify which patients have a "clinically important brain injury," for which the current "Gold Standard" is a positive result on the computed tomography (CT) scan [10]. Attempts to develop "clinical decision rules" for obtaining a CT scan in the emergency department (ED), [e.g., New Orleans Criteria (NOC) and the Canadian CT Head Rule (CCHR)], have resulted in extremely high sensitivity for clinically significant brain injury, (95%–100%), with extremely low specificities (∼12%–20%) [11], [12]. Furthermore, the growing awareness of the possible radiation risks associated with CT scans further complicates its use in the diagnosis process and highlights the shortcomings of current decision rules. In contrast, the large majority of TBI patients who are CT negative (CT-) or for whom a CT is not medically indicated (mTBI or concussion), may have significant abnormalities in brain function for which no "Gold Standard" currently exists, despite attempts to establish guidelines to identify and determine the severity of such concussive injury [13].

Mild TBI patients have been separated from moderate/severe TBI and from normal with high accuracy using linear discriminant functions built on variables extracted from 19-lead QEEG

Manuscript received December 01, 2011; revised April 26, 2012; accepted June 15, 2012. Date of publication July 26, 2012; date of current version November 02, 2012. This work was supported by BrainScope, Inc., Bethesda, MD, USA.

L. S. Prichep is with the Brain Research Laboratories, Department of Psychiatry, New York University School of Medicine, New York, NY 10016 USA (e-mail: leslie.prichep@nyumc.org).

A. Jacquin, Julie Filipenko, S. G. Dastdar, and N. Rothman are with BrainScope, Inc., Bethesda, MD 20814 USA (e-mail: arnaud.jacquin@brainscope.com; julie.filipenko@brainscope.com; sam.dastidar@brainscope.com; neil.rothman@brainscope.com).

S. Zabele is with Raytheon, BBN, Cambridge, MA 02138 USA (e-mail: szabele@bbn.com).

A. Vodenčarević is at the International Graduate School Dynamic Intelligent Systems, 33102 Paderborn, Germany (e-mail: asmirv@gmail.com).

Digital Object Identifier 10.1109/TNSRE.2012.2206609

[6], [14], [15]. Features of brain function which contributed most to these separations were those that reflect changes in power and synchrony relationships between brain regions, reflecting disturbances in "connectivity." It is important to note that such measures are the ones which have been suggested to correlate with the abnormalities reported using diffusion tensor imaging (DTI) in diffuse or traumatic axonal injury (DAI or TAI), suggested to be of etiological significance in concussion [16]–[18]. Cao *et al.* have shown that classical quantitative QEEG variables extracted from the EEG power spectrum could successfully detect functional deficits following concussion in a study of 61 athletes [19]. These researchers all used EEG data collected from the full set of 19 leads standardly positioned.

Although previous QEEG findings in mTBI are suggestive, a method for classification which is based on 19 lead EEG data is clinically impractical, since resources to obtain such data in the acute setting are typically not available or too cumbersome to employ. The EEG input to the algorithms described in this paper was therefore restricted to a more clinically viable limited montage of data acquired from only frontal forehead locations. The use of this limited montage is supported by considerable evidence of the maximum vulnerability of frontal regions of the brain to TBI, regardless of the direction of the forces, of where the trauma was sustained, or of the etiology of the injury [13], [16], [20], [21]. Furthermore, recent evidence from DTI studies have reported abnormalities in frontal fiber tracts in mTBI [13], and demonstrated that frontal tract abnormalities were most highly correlated with EEG phase synchrony between frontal regions [17]. Such evidence supports the unique importance of these brain regions to the disruption of connections thought to be fundamental to concussive injury, and suggests that recordings made over the frontal regions alone should be sufficient in obtaining high sensitivity to mTBI.

Important to this study was the use of statistical classification in the separation of mTBI subjects into clinically meaningful categories. A generic multiclass classification task can be thought of as a mathematical function (linear or nonlinear) which uses as its input a vector of computed quantitative *features* of an object (also called *observations* or *patterns*) and produces as output, a label which assigns the object to a specific *category* (or *class*). Suppose that we want to build a classification system (also called *classifier function*) which separates objects of interest into $C$ categories $\mathrm{c}_1, \mathrm{c}_2, \ldots, \mathrm{c}_\mathrm{C}$. Suppose in addition, that for each incoming object, $X_i$, a set of K quantitative features: $x_1, x_2, \ldots, x_k$ are computed. The classifier function $F$ should be designed such that it uniquely assigns a category label $c_j$ to each incoming object $X_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,k})$. This can be written symbolically as

$$F(x_{i,1}, x_{i,2}, \ldots, x_{i,k}) = c_j. \tag{1}$$

When there are two categories, the classifier function is often referred to as a *binary classifier function*. Duda *et al.*, [22] noted that "the degree of difficulty in a classification problem depends on the variability in the features values for objects in the same category relative to the difference in feature values for objects in different categories." In any nontrivial classification problem, the distributions of features for objects in two or more categories will significantly overlap. Multiclass classifiers (with $C$ categories) are usually designed by combining the outputs of up to $C$ binary classifier functions [23]. A binary classifier function $F$ for the two classes $c_1$ and $c_2$ is usually derived in a straightforward way from a discriminant function $g$ which assigns a value $g_i$ to an incoming object $X_i$. This derivation is done according to the following assignment rule:

$$\begin{aligned} \text{If } g(X_i) \leq T \text{ then } F(X_i) &= c_1 \\ \text{otherwise } F(X_i) &= c_2 \end{aligned} \tag{2}$$

where $T$ indicates a numeric *classification threshold*. Moving this threshold in overlapping distributions results in a trade-off between sensitivity and specificity, emphasizing the importance of considering the stratification of risk (i.e., false positive versus false negative) as appropriate for the specific application of the classifier function. To classify subjects into one of several categories, multiple approaches can be used. A single three-class classifier identifies the category with the highest likelihood of membership for each subject, but this approach is less effective when there is considerable overlap between the categories (as is the case in our TBI categories). A set of three binary classifiers (i.e., 1 versus 2, 2 versus 3, and 3 versus 1) can be used with a voting strategy to combine the three results and determine the appropriate category for the subject. However, this approach often results in confounding results or disagreement between two classifiers that must be resolved. Furthermore, as often occurs with biological signals, the subject classes are not well separated (highly overlapping), potentially causing lack of reliability and diminished utility of the classification method. In this paper, we describe a method that enables such a subject population, patients who have suffered a head injury, to be classified into clinically useful categories using a multidimensional set of QEEG measures, informed data reduction, and a sequence of binary classifiers.

## II. Methods Data Acquisition

### A. Subjects

Data was collected at 13 EDs across the US, with approval from local Institutional Review Boards (IRBs). Subjects were a convenience sample ($n = 633$) of males (70%) and females meeting inclusion/exclusion criteria described below. In 499 of these cases a second sample of non-overlapping data was selected from the same recording to be used as a stability sample. All subjects signed written informed consent.

*Inclusion Criteria:* Males and females between the ages of 15 and 80, who suffered a closed head injury and with a Glasgow Coma Scale score (GCS, [24]) above 8, with or without loss of consciousness (LOC) or traumatic amnesia and with symptoms of TBI. *Inclusion criteria for "normal controls"*: 1) ED patients under duress 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 injury; and 2) subjects who sustained head injury but had no altered mental status (AMS), no LOC, no amnesia and no significant symptoms related to head injury upon presentation.

*Exclusion Criteria:* Subjects with scalp or skull abnormalities or whose clinical condition, such as head trauma, will not allow placement of the electrodes; intoxication in those

TABLE I
AGE DISTRIBUTION BY CATEGORY

| | AGE RANGE (years) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| CAT.* | <18 | 18-30 | 31-40 | 41-50 | 51-60 | 61-70 | 71-80 | >80 |
| **1** | 0 | 111 | 31 | 32 | 27 | 15 | 2 | 0 |
| **2** | 35 | 124 | 44 | 48 | 28 | 17 | 9 | 1 |
| **3** | 0 | 15 | 14 | 13 | 21 | 15 | 24 | 7 |

*Total for category **1**: 218, **2**: 306, **3**: 109

obtunded to the point where they could not participate in the study. In addition, the following patients were excluded: those with advanced Dementias, Parkinson's disease, chronic drug or alcohol dependence, known seizure disorder, mental retardation, or those taking daily prescribed medication for a known psychiatric disorder.

In a triage setting, complicating factors such as drugs, alcohol, fatigue, pain, etc., are present and can affect the clinical presentation of a patient and the characteristics of their EEG, and therefore need to be considered to enable accurate classification. The strategy adopted in our investigation was to not attempt to exclude these factors, but rather include them in all subject groups. By doing this, they are eliminated as differentiating factors among groups, and features sensitive to these factors are not selected for the two classifiers that enable separation of subjects into categories.

### B. Classification of TBI

Since the statistical classification approach described in this paper belongs to the category of *supervised learning methods*, it required the assignment of each subject in the training database to a category specific to severity of TBI. Since there is no established "gold standard" for mTBI or agreed upon definitions of categories within the spectrum of TBI, we consulted with emergency medicine and sports medicine physicians in conjunction with published guidelines [25], [26], to define three clinically relevant categories for subjects suspected of a traumatically induced structural and/or functional brain injury.

1) *Category 1 ("normal controls")*: Absence of acute traumatic structural brain injury (visible on CT scan or CT deemed unwarranted) and no evidence of functional brain injury.
2) *Category 2*: Absence of acute traumatic structural brain injury (visible on CT scan or CT deemed unwarranted), but evidence of functional brain injury (based on validated symptom and neurocognitive assessments, as determined by the expert group of physicians).
3) *Category 3*: Presence of acute traumatic structural brain injury (visible on CT scan, CT+).

These categories, therefore, provide both the basis for training the classification algorithms and a standard against which their performance can be evaluated in a prospective study or with an independent test group. Table I shows the total number of subjects in each category and their age distribution by decade from 15 to $> 80$ years. It should be noted that 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.

### C. EEG Data Acquisition and Artifact Detection

*EEG Protocol:* Ten minutes of eyes closed resting EEG was recorded from the following locations of the expanded International 10–20 Electrode Placement System: Fp1, Fp2, F7, F8, AFz, A1, and A2. FPz was used as ground electrode. Signals were remontaged to linked ears per standard QEEG practice, and consistent with the normative data described below. A sampling rate of 8 kHz[1] was used and the data was subsequently down-sampled to 100 Hz for processing.

*Automatic Artifact Detection:* The classification system described in this work included a fully-automated artifact detection module (comprised of a set of algorithms functionally described below). Approximately one to 2 min of artifact-free EEG data (representing 24–48 epochs of length 2.56 s) were selected from the recording using the artifact detection algorithms described briefly below. It was required that each epoch be created from continuous EEG data. The 2.56 s epoch corresponds to an FFT size of 256 points used to compute estimates of the power spectra of the subject's EEG.

The artifact detection module identifies the following types of artifacts, briefly summarized below.

*Type 1.a: Vertical eye movement (VEM)*. Detection of a vertical eye movement (VEM) is performed by locating large excursions ("peaks") beyond a given threshold on the Fp1 and Fp2 leads. Since both eyes move in unison, the algorithm ensures that only such excursions which occur *concurrently* and *in the same direction* (same polarity of the peaks) are identified as vertical eye movements.

*Type 1.b: Horizontal/Lateral Eye Movement (HEM/LEM)*. Slow lateral eye movements (HEMs) produce waveforms of that have opposite polarity at F7 and F8. They are characterized by a low-frequency ($\leq$3 Hz) "scissor-pattern" best seen on these two electrodes. Thresholded differences between these leads filtered in the band of interest (0.5–3 Hz) are identified as artifact.

*Type 2: Patient Cable (or Electrode) Movement (PCM)*. This artifact typically produces extremely large slow waves in the EEG traces. It is detected by identifying excessively large EEG magnitudes (also called "over-range"), using an amplitude threshold above that thought to occur normally in brain activity in any of the five recorded frontal EEG channels.

*Type 3: Impulse artifacts (IMP)*. This algorithm looks at high-frequency variations of signal amplitude in each subepoch. Within each window examined (100 ms width), the value $(\max - \min)$ is computed and triggers an IMP artifact detection when it exceeds a given threshold. Care is also taken to eliminate "sharp alpha" as a trigger of IMP, using an additional filter.

*Type 4: Muscle activity (EMG)*. This artifact is characterized by high-frequency signals (above 20 Hz) occurring in bursts of variable duration, as reported in the literature [27]. Its effect on the power spectrum of the EEG is to modify its general morphology towards a "flat spectrum," which is a characteristic of the power spectrum of White

[1]8 kHz sampling was used to enable functionality to collect brainstem evoked responses not used in this work which therefore used only the 100 Hz data.

Gaussian Noise. As such, this artifact can be detected by comparing the power in two neighboring EEG frequency bands.

*Type 5: Significantly Low Amplitude Signal (SLAS)*. This artifact is meant to capture extremely low-amplitude EEG signals (at all frequencies) which occur, for example, when the brain is in Burst Suppression mode; a condition which can occur (but should be avoided) during anesthesia. This artifact can be detected by looking for signal epochs with mean-square energy below a threshold.

*Type 6: Atypical Electrical Activity Pattern (AEAP)*. This artifact type is meant to detect unusual patterns of activity in the EEG such as those which occur in the EEG of epileptic subjects during a convulsive or nonconvulsive seizure. The algorithm is sensitive to spike-wave complexes occurring in bursts over several hundred milliseconds. It uses a combination of wavelet analysis and fractal dimension computation and was presented in detail in [28].

Note that out of these artifact types, two are nonphysiological (type 2, type 3), three are physiological but are not brain-generated (type 1a–1b, type 4) and two are brain-generated (type 5, type 6). All of these artifacts reflect nonstationarity of the noisy EEG signal. Examples of artifacts of these types, as they appear on a 19-lead electrode montage, can be found in a standard EEG reference text [29]. When an artifact is identified in any lead, data from all leads are removed from that time period. Thresholds referred to in the above summaries were determined in interactive studies with expert visual analysis of the records, and were based on good performance on the match between technologist-based and automatic detection of the artifact type for the training dataset. It is important to note that comparison of the automatic artifactor with visual editing of an EEG expert resulted in a high percentage of data overlap (87.6%) [30], which is significantly higher than the average inter-rater agreement reported in the literature between visual editors [31], thereby validating the assumption that automatic editing of EEG performs at least as well as an expert EEG technologist.

## III. Construction of Database of Quantitative EEG Features

The classification algorithms reported on here derive their performance from our diverse set of QEEG features, both linear and nonlinear. Features include both traditional features derived from estimates of power spectra [32] computed in the conventional frequency bands as well as nontraditional quantitative features computed using state-of-the-art signal processing methods. The nontraditional features are reviewed in detail elsewhere by Thakor *et al.* [33] and Sakkalis *et al.* [34]. It is also noted that since there is no single unique solution to the classification problem posed, and considering the high correlation between QEEG variables in the reduced space of the frontal regions, we used an expanded set of measures considered to be descriptive of perturbations of the QEEG in clinical populations. Below, we give the formulas used to derive the features that are part of our feature set.

### A. Univariate Features

*1) Age-Regression and $z$-Transform of Univariate Features:* The EEG of a normally functioning person in the resting or ground state of the brain is regulated by an anatomically extensive, genetically based neurophysiologic homeostatic system which changes predictably as a function of age. The relationship with age has been demonstrated to be well described by a set of regression equations across ages 6–90 years, which have been published [32], [35], [36], replicated in a series of international peer reviewed publications and demonstrated to enhance clinical utility of QEEG [4]. The sample of normal/control subjects used in these equations is referred to as the *norming group*. In order to remove age as a factor in the development of the classification algorithms, the development process includes an *age regression* step which is performed in order to eliminate the (normal) influence of age on any univariate variable computed from the EEG. A model is assumed in which a raw variable (transformed) varies approximately linearly with the $\log_{10}$ of the patient age expressed in years. For each QEEG variable $v_i$ the age-regression parameter $a_i$ is therefore computed (offline and only once on the norming group) as the slope of the best-fitting straight line for the cloud of points $(\log_{10}(\mathrm{SubjectAge}), v_i)$. The age-regressed variable will consequently have a linear fit of slope zero with the variable $\log_{10}(\mathrm{SubjectAge})$, which results for this variable in a very small correlation with age

$$y_i = v_i - \log 10(\mathrm{SubjectAge}) \cdot a_i. \tag{3}$$

The *$z$-transform* is a standard statistical transform which normalizes the distributions of QEEG variables to an approximately Gaussian distribution. Z-transformed values of univariate variables are obtained by the equation

$$z_i = (y_i - m_i)/\sigma_i \tag{4}$$

where $m_i$ and $\sigma_i$ denote, respectively, the mean and standard deviation of the age-regressed variable $y_i$. These two scalar parameters are also computed offline for the *norming group*. After this process, all features will have a mean $= 0$ and standard deviation $= 1$ for the norming group and will be in units of standard deviations, i.e., the feature will be expressed on a common metric of probability. This facilitates the use of combinations of features in discriminant functions without the complication of different sets of units, as well as the creation of multivariate features.

Examples of two specific variables $v_i$ and $y_i$ with their corresponding best linear fits with are shown in Figs. 1 and 2. In these figures, we illustrate the age-regression process on two sample QEEG variables: 'BRF1F2A' ("Bipolar Relative Power variable in Alpha band, between Fp1 and Fp2") and 'BFrF7ZS' ("Bipolar Fractal Dimension in S band, between F7 and AFz.").

The top panels of Figs. 1 and 2 show the scatter plot of $(\log_{10}(\mathrm{SubjectAge}), v_i)$, i.e., of the variable prior to age-regression. Note that in each figure, the slope of the best linear fit is not zero indicating the presence of a correlation of this variable with age. In contrast, the bottom panels of these figures show the scatter plot of $(\log_{10}(\mathrm{SubjectAge}), z_i)$, i.e., of the variable after age-regression and z-transform. Note that in

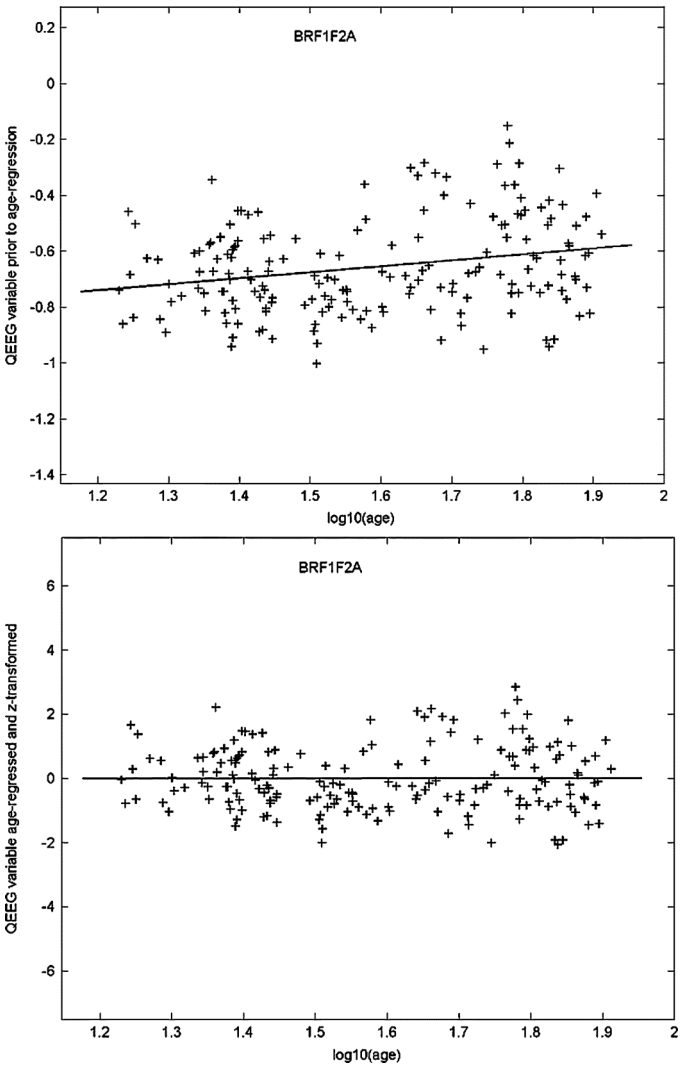


Fig. 1. Top panel shows scatter plot of $(\log_{10}(\mathrm{SubjectAge}), v_i)$ for univariate variable "BRF1F2A" (not age-regressed), for the 180 norming subjects. The straight line indicates the best linear fit. Bottom panel shows scatter plot of $(\log_{10}(\mathrm{SubjectAge}), z_i)$ for univariate variable "BRF1F2A" (age-regressed and z-transformed), for the 180 norming subjects. The (horizontal) straight line indicates the best linear fit after age-regression.

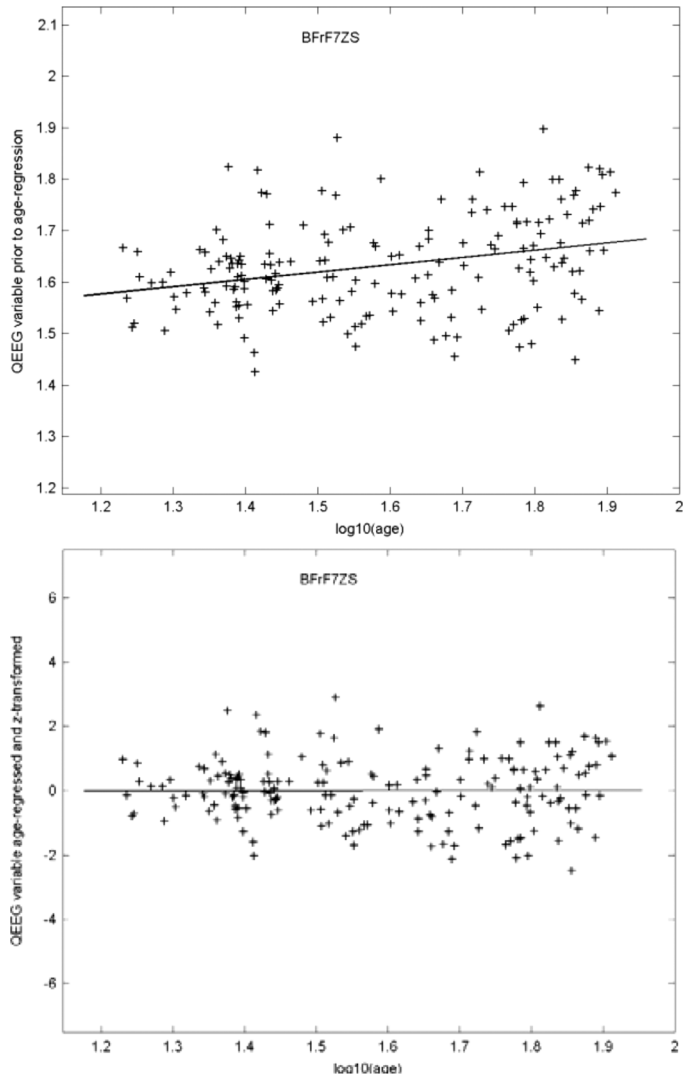


Fig. 2. Top panel shows scatter plot of $(\log_{10}(\mathrm{SubjectAge}), v_i)$ for univariate variable "BFrF7ZS" (not age-regressed), for the 180 norming subjects. The straight line indicates the best linear fit. Bottom shows panel scatter plot of $(\log_{10}(\mathrm{SubjectAge}), z_i)$ for univariate variable "BRF1F2A" (age-regressed and z-transformed), for the 180 norming subjects. The (horizontal) straight line indicates the best linear fit after age-regression.

each figure, the slope of the best linear fit is zero indicating that the correlation of this age-regressed variable with age is insignificant.

*2) Traditional Univariate Features:* These features are computed in 10 frequency bands: Delta1 (0.5–1.5 Hz), Delta (1.5–3.5 Hz), Theta (3.5–7.5 Hz), Alpha (7.5–12.5 Hz), Alpha1 (7.5–10 Hz), Alpha2 (10–12.5 Hz), Beta (12.5–25 Hz), Beta2 (25–35 Hz), Gamma (35–50 Hz), and Total (1.5–25 Hz). Traditional QEEG features in our set consisted of monopolar absolute and relative power, bipolar absolute and relative power, monopolar and bipolar mean frequency, monopolar and bipolar inter- and intra-hemispheric power asymmetry, monopolar and bipolar inter- and intra-hemispheric coherence. These features are derived from the EEG power spectrum, cross-spectrum and covariance matrix and are described in detail elsewhere [32], [36]. Following the computation of these "raw features" a log-based transform was applied as described in the literature [32], [35], [37] to improve the Gaussianity (or normality) of these features prior to further statistical computations.

*3) Nontraditional Univariate Features:* In addition to traditional univariate features, we also use *chaotic/fractal measures* (fractal dimension and scale-free activity), *information theory-based measures* (entropy and wavelet entropy), and *functional connectivity measures* (phase lag and phase synchrony). Measures of the first two types are computed on the full-band signal only. The replicability criterion presented in Section IV-A serves, for these chaotic measures, as a test for quasi-stationarity. These features add to the dimensionality of the total feature set and expand importantly for TBI, in the domain of connectivity and other measures of disturbances of the EEG signal beyond the frequency domain. Further details of the computation of these measures are provided in Appendix A.

*a) Fractal Dimension Measures:* This measure evaluates the global complexity of the brain electrical activity at each elec-

trode location across the total spectrum. The fractal dimension of a 1-D signal is the fractal dimension of the curve formed by the plot of that signal. It is a nonlinear mathematical quantity (taking values between 1 and 2), which reflects to what degree the curve fills 2-D space as one zooms down to finer and finer scales.

*b) Scale-Free Brain Activity:* Like the fractal dimension this measure evaluates the global complexity of the brain electrical activity at each electrode location. The power spectrum of spontaneous brain activity typically follows a $1/\mathrm{f}^{\beta}$ law over the frequency range 1–25 Hz and is therefore approximately linear when plotted on a log-log scale. This behavior of the power spectrum is typical of fractal processes. The parameter $\beta$ is usually referred to as *Scale-Free Brain Activity* [38].

*c) Information Theory-Based Measures:* Four types of these measures are computed: traditional Shannon entropy, Tsallis entropy, wavelet entropy and relative wavelet entropy. These entropy measures evaluate, each in a slightly different way (see Appendix A), the degree of order/disorder of the brain electrical activity at each electrode location.

*d) Functional Connectivity Measures (Phase Lag and Phase Synchrony):* These are measures of functional connectivity and evaluate the relationship between and among brain regions. Phase lag was computed, in each frequency band, as the average phase delay between signals at two electrodes using the normalized cross-spectrum [39]. Phase synchrony was computed from a time-frequency representation of the EEG (RID-Rihaczek distribution) and measures the degree of phase locking between two electrode locations in each frequency band [17]. It is of note that we also computed mutual information (MI) and cross mutual information (CMI) measures, but they were not found to meet our entry criteria or were not selected by the classification methods and so are not described in more detail.

### B. Multivariate Features

The set of QEEG variables also includes sets of *multivariate anterior* features which are non-linear functions of selected groups of the univariate variables described above. These multivariate features are computed "across regions," for any given QEEG frequency band. Therefore, they correspond to a larger, less focal region and help in capturing the functional performance of this brain area treated as a "system." Currently, multivariate features are derived for combinations of regions which include: all regions, all left hemisphere regions, all right hemisphere regions, prefrontal regions, frontotemporal regions, and cross-frontal regions.

Multivariate variables are computed as a cube root of sums of squares of z-transformed, age-regressed univariate variables as follows:

$$Z_i = \left(z_1^2 + z_2^2 + z_3^2 + \ldots z_k^2\right)^{1/3} \tag{5}$$

where $k$ denotes the number of univariate variables $(z_1,\ z_2,\ \ldots z_k)$ included in the computation of the multivariate. Since the univariates are approximately Gaussian, the sum-of-square has a Chi distribution (i.e., is positively skewed). This skew is reduced by the application of the cube root.

A multivariate variable $Z_i$ computed by (5) may no longer have mean of zero and standard deviation of 1 for the norming group and may also show a dependency on age. Thus, a second stage of age-regression and z-transform is applied to each multivariate variable as follows:

$$z_i = (Z_i - \log_{10}(\text{Subject Age}) \cdot b_i - M_i)/S_i \tag{6}$$

where $b_i$ is the age-regression parameter (determined for the norming group as described above) and $M_i$ and $S_i$ are the mean and standard deviation of the age-regressed variable $Y_i = Z_i - \log_{10}(\text{SubjectAge}) \cdot b_i$.

## IV. Informed Data Reduction and Feature Selection Methods

### A. Informed Data Reduction

Our complete pool of quantitative features consists of a total of 1536 features (1215 univariate, 321 multivariate) extracted from each recording. With such a large number of features, it is important to apply data reduction methods prior to selecting features for the classifiers. Conventional methods of data reduction reported in the scientific literature include t-tests and ANOVAs, used to identify variables which are significantly related to dependent variables of interest [40]. Variables which maximize adjusted multiple correlation coefficients between QEEG and dependent variables, minimizing the residual sum of squares (RSS), are selected. While these methods are of some limited use, they do not systematically address important considerations of adequacy of feature selection, and may be inappropriate to the construction of nonlinear (e.g., quadratic) discriminant functions where means differences may no longer be of primary importance. In this study, data reduction is advanced by using an "informed" approach to the variable/feature pool, including the following requirements for variables to be included in the set available to the classifier construction methods.

1) *Replicability:* While there is a literature which attests to the stability and replicability of classical QEEG features across time when there is no known change in brain state [41], the replicability of each quantitative feature in this study was evaluated separately for short-term replicability, where only those features above a certain replicability level were candidates for input to algorithm development. To test for replicability within this dataset, QEEG features were extracted from a large population of normal and abnormal subjects $(n = 1400)$ from a first set of 48 clean epochs (1–2 min) and then computed again from a second set of 24–48 clean epochs (within the same recording and immediately following the first, when available). For each QEEG feature $\mathrm{x_k}$ (z-score), we compute the mean $m_k$ and standard deviation $\sigma_k$ across subjects, of the magnitude of the difference between the z-score and its replication, i.e.,

$$m_k = \text{mean}\left(|x_{k,1} - x_{k,2}|\right) \tag{7a}$$

$$\sigma_k = \text{std}\left(|x_{k,1} - x_{k,2}|\right) \tag{7b}$$

We then compute the "variability" $v_k$ of the features as

$$v_k = m_k + 3\sigma_k. \tag{7c}$$

A "replicability score" (either 0 or 1) is then assigned to each feature by comparing the variability to a fixed threshold $T$ where if $v_k \geq T$ the replicability score $= 0$, or else replicability score $= 1$.

A value of $\mathrm{T} = 1.8$ was selected for the threshold using empirical techniques to balance replicability requirements with the need to preserve a sufficient number of features in the variable database. By requiring replicability in all features in the database, we ensure that we are representing the quasi-stationary characteristics of the EEG which are most important in characterizing brain activity in this application.

2) *Separability:* It was hypothesized that selecting individual features that are unequivocally able to separate the two groups in a binary classification task will improve the overall multivariate performance of the classifier function in terms of sensitivity and specificity. Therefore, a two-sample Kolmogorov–Smirnov test was employed to test the null hypothesis that the values for an individual feature for the two groups are from the same distribution. The alternative hypothesis is that they are from different distributions. The significance level was selected to be 0.15 so as not to remove too many features. Any feature that did not pass this test (i.e., null hypothesis was not rejected) was removed from the feature pool available for the construction of the classifier algorithm.
3) *Homogeneity:* It was also hypothesized that the consistency of classification performance could be improved by ensuring that we select those features that have similar distributions within each group considered separately. Therefore, the data from each group were randomly divided into two sets and a two-sample Kolmogorov–Smirnov test was employed to test the null hypothesis that the values for an individual feature for the two groups are from the same continuous distribution. This process was repeated. The significance level was selected to be 0.05. Any feature that did not pass this test (i.e., null hypothesis was rejected) for any one of the two repetitions was removed from the feature pool available to the classifier construction algorithm.
4) *Normal means value:* As described above, all features are transformed to z-scores, expressed in standard deviation units of the norming group. Since all features for the norming group population have a mean of 0 and a standard deviation of 1 (by definition), it may be expected that the normal controls in a training population will also have approximately the same mean and standard deviation. Thus, for a feature to be considered a candidate for the classification algorithm, it was required that it have a mean value less than 1.0 in the normal training population.
5) *Absolute means difference:* Following the same clinical logic as for 4 above, the characteristics of the mean values of a feature in the "normal" training population should be closer to that of the normative population than the "abnormal" training population with a structural head injury (category 3). This condition is implemented as a rule whereby all features for which the absolute value of the mean for the abnormal population must exceed that for the normal population by at least 0.01 to be included in the available reduced feature pool for classifier selection. In this way variables were eliminated that had spurious differences between groups, i.e., where the normal group had the more abnormal value.
6) *6. Neurophysiology-based exclusion:* To avoid potential contamination of the feature pool, certain variables were excluded *a priori* when: 1) the literature shows them to be inherently nonreplicable, 2) there existed an inadequate norming sample, or 3) the feature may be affected by remaining artifacts (e.g., features for the delta1 band). In this way, the potential for confounding factors not related to the classification issue of interest would be minimized.

The use of such a method to help prune the size of available features from a pool of over 1500 to a range (300–500) that can be reasonably assessed using current generation computational platforms is a major strength of this approach. Combining these steps in the process of data reduction enriches the pool of variables that are entered into Section IV-B. It is noted that the thresholds used in the steps described above are determined with respect to the size of the feature set and number of subjects and would be modified accordingly for different numbers of features and subjects, such that a sufficient number of features remain.

### B. Feature Selection Methods

To obtain optimal classification performance, two feature selection methodologies were investigated, genetic algorithm (GA) and deterministic feature selection (SFP). Both methodologies are based on different approaches but have a common goal, i.e., to create increasing performance of the classifier functions at each iteration of the process. These two methods are described below.

1) *Genetic algorithm and modified Random Mutation Hill Climbing optimization:* Following the informed data reduction steps described above, a large number of features still remain in the reduced feature pool (denoted by $M$) for use in each classifier, which typically leads to a very large value for the total number of distinct sets of $K$ features taken from a total pool of $M$ features (which would corresponds to the size of the search space in an exhaustive search). In order to select near-optimal subsets of $K$ features for a classifier function, a state-of-the-art GA was used. GAs are special types of a class of search algorithms called *evolutionary algorithms* that are used for solving a wide range of optimization and search problems [22], [42]. These algorithms mimic the biological process of *evolution* or, more descriptively, the process of *survival of the fittest* (see for example [22, Sec. 7.5]). They have been used extensively in machine learning applications and are particularly good at solving classes of problems involving a search through a very large space of possible solutions, including feature selection problems [43]–[46]. This approach was uniquely suited to the variable selection and classifier problem faced in this study, although not previously used in the literature to design classifiers of brain function using subsets of QEEG features.
   - Fig. 3 shows a block diagram of the GA search technique applied to the problem of selecting a "best set" of

$K$ features among $M$ ($M \gg K$) in view of performing a classification task using a discriminant function (DF) built from the $K$ features. First, any subset (list) of $K$ features from our total pool of $M$ features is represented as a binary bit string of length $M$, which we refer to as a "chromosome" (in accordance with GA research literature), where $K$ bits are set to one and all other bits are set to zero. Each bit corresponds to a single feature. If the bit value is one (respectively zero), the corresponding feature is present (respectively absent) in that particular subset of features. According to this representation, the "genes" of the chromosome are the individual bits.

- GAs require the selection of an *objective function* [also called *figure of merit (FOM)*] which should reflect the classification performance of any chromosome. In this model, the area under the curve (AUC) for the receiver operator characteristic was used as the objective function[2] since a high AUC (close to 1) is associated with high sensitivity/specificity of the DF [47]. Specifically, the AUC was computed using a ten-fold cross-validation method. During the initialization phase, the GA-based classifier builder produces $P$ random binary bit strings (this initial chromosome population of size $P$ is referred to as "generation 0"), where each chromosome represents a randomly selected subset of $K$ features. By application of the standard *GA operators* (*ranking, selection, recombination*, (*crossover*), *mutation* and *reinsertion*), the initial population of chromosomes "evolves" through $G$ generations towards a population with better overall fitness. The search for "best discriminant functions" with $K$ variables can therefore be seen as the search for "best chromosomes" (in the sense of having a high objective function value) in each generation, containing $K$ bits set to one.
- The parameters of the GA algorithm were the following: 1) population size ($P$): 50; 2) number of generations ($G$): 50; 3) recombination probability: 0.6; 40 mutation probability: 0.003. These parameters are standard for GA algorithms, and it was found that the final results from GA runs were robust to modifications of these parameters.
- After $G$ generations, the overall best chromosome is further improved through a local random search using a *modified Random Mutation Hill Climbing* algorithm (mRMHC). This last step is performed because while the GA will typically produce a "highly fit" generation of chromosomes after a number of iterations, it does not guarantee that a local optimum is reached. This latter algorithm simply performs a local search around the best chromosome found among all generations, "modified" in such a way to keep the number of features in the chromosome constant. This is done by slightly modifying the classical RMHC algorithm (see for example [42, p. 129]) so as to randomly flip two bits (of different values) instead of a single one, thereby repeatedly substituting a single feature for another.
- 2) *Deterministic feature selection* (the "Simple Feature Picker $k$ at a Time," SFPx$k$): The Simple Feature Picker (one at a time) (acronym: SFPx1), was so named by the authors because the deterministic process is conceptually simple. It is a stepwise deterministic process that selects one feature at a time based on a FOM. In this case, the FOM was the average of the AUC of the classifier receiver operating characteristic (ROC) curve computed using a ten-fold cross-validation method (described below). The first feature used in the classifier is selected from the available feature pool by finding the one that yields the largest average AUC based only on that single feature. The next feature added to the classifier from those remaining in the feature pool is the one that yields the largest increase in the AUC for a classifier consisting of the pairing of that feature and the previously selected feature. The process continues until a predetermined number of features are selected or there are no more features available that increase the AUC. One of the distinct problems with using the SFPx1 algorithm, particularly with large numbers of very similar features and for classification problems with very similar groups, is that it can easily latch on to a local maxima early in the process. This tendency often prevents it from finding what are in some cases significantly better solutions as more variables are considered later into the process. To address this shortcoming, we modified the basic SFPx1 algorithm to retain the best $k$ solutions ($k > 1$) at each iteration, rather than the single best solution. This enhanced algorithm, referred to as the SFPx$k$ algorithm, considers each of the $k$ solutions as a starting point for the next iteration, and chooses the subsequent set of $k$ best solutions from the feature pool. To find the best trade-off between run time and solution performance, we ran a series of experiments for values of $k$ between 2 and 10. We found $k = 5$ to be a reasonable choice. The mRMHC process was also applied to the SFPx$k$ solutions to further enhance these solutions.

### C. Rules to Prevent Over-Fitting

An important consideration in the development of an effective classifier is that it not be over-fitted to the training data, resulting in a loss of generalization to the broader population. Over-fitting typically results in very good performance on the training data, but poor performance on an independent test set. To guard against it, we imposed a limit on the number of features used in constructing each classifier. Statistical methods to help determine the optimal subject to variable ratios have been suggested, including a minimum subject-to-variable ratio of 10:1 for a linear discriminant model [48]. For a quadratic discriminant model, the number of variables selected, $N$, should be such that $N(N+3)/4$ does not exceed the number of samples in the smallest training category [48]. In the case of our 3 versus 2,1 quadratic discriminant function (QDF), the smallest group is Category 3 subjects (109 subjects) and therefore the maximum number of variables which we allow for this discriminant is 19 since $(19 \times 22)/4 < 109$.

[2]A cost penalty was added to the AUC which increases linearly as the number of features in the classifier departs from a *base number* of features. The base number was selected such that it meets the maximum number of features rule to prevent over-fitting.

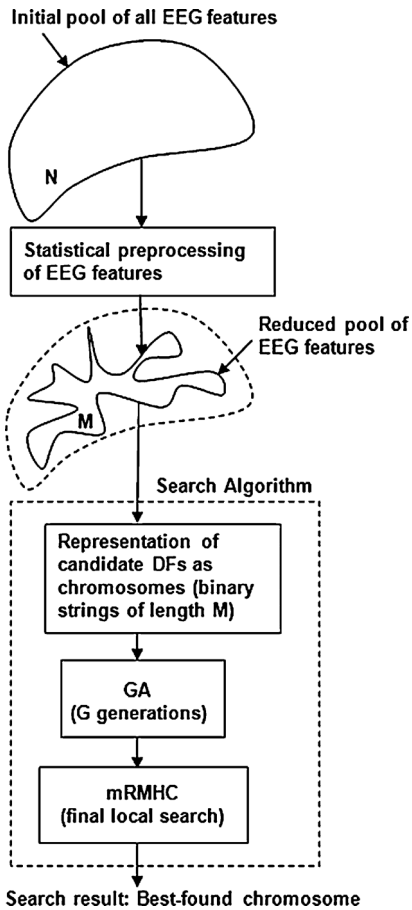


Fig. 3. Block diagram of exemplar linear discriminant function (LDF) building system involving: 1) statistical preprocessing of initial pool of quantitative brain activity features, 2) the representation of candidate LDF solutions as "chromosomes," 3) evolutionary search strategies (GA with mRMHC algorithm).

Beyond simply limiting the number of features selected, an attempt was also made to ensure that each selected feature provided a significant and independent contribution to the information content of the classifier function. This rule was implemented during the feature selection process (both GA and SFP) whereby only features that were highly correlated (i.e., Pearson correlation coefficient greater than 0.95) with any existing feature in the feature list could not be selected. This directly reduced the possibility of over-fitting due to redundant information.

### D. Cross-Validation

As with many classifier development efforts involving the use of human subjects, we too were faced with the problem of having very limited sized populations. In order to use all available data for both testing and training in a way that was statistically validated, cross-validation methods were used. Cross-validation involves iteratively dividing a population into nonoverlapping test and training groups, where, at each iteration, the training group is used to develop a classifier and the test group is used to evaluate the performance of that classifier. The process repeated until all data has been used for both testing and training, with an aggregate performance measure constructed by combining the individual test performance measures obtained at each iteration.

In this work, we use cross-validation both during the feature selection phase of binary classifier development and for evaluating classifier performance once a reduced pool of features has been identified. For the feature selection phase, as described above we use *ten-fold* cross validation [49], ensuring that each of the three categories of subjects is equally represented in each fold. To evaluate classifier performance after the subset of features (selected from the reduced pool of features) has been finalized (using GA or SFP), a *leave one out* (LOO) cross-validation was used. LOO cross-validation can be viewed as the limiting case of an N-fold cross-validation in that at each iteration only one subject is used for testing while the remaining population is used to train the classifier. Hence, if there are $N$ total subjects across the combined categories, we develop and test $N$ different classifiers. We again compute an aggregate performance measure by combining the individual performance measures from the $N$ different classifiers. The use of these two different cross-validation methods is intended to provide both conservative and reliable predictions of future classifier performance.

### E. ROC Curves and Performance Metrics

In order to obtain a statistical description of the performance of a given two-class classifier function, it is useful to compute and plot the ROC curve. The ROC curve illustrates, for any choice of a discriminant output *threshold*, the performance [*sensitivity, specificity, positive predictive value (PPV)*, and *negative predictive value (NPV)]* which can be expected of the binary classification algorithm. For definitions of ROC, sensitivity, specificity, PPV, and NPV in the context of our discriminants, the reader is referred to Appendix B. A useful global scalar measure of performance of a binary classifier is the AUC.

### F. Sequence of Binary Classifiers

The algorithm used in this work employs two binary discriminant functions (DF), constructed from the reduced pool of features extracted from the EEG, defined as follows: 1) structural injury versus nonstructural injury/normal—Category 3 versus Categories 1 and 2 (DF_3v21) and 2) abnormal versus normal-Categories 2 and 3 versus Category 1 (DF_32v1).

To classify a subject, we apply these classifiers sequentially, following simple category assignment logic: If subject X is classified as "3" by binary classifier DF_3v21, assign the subject to category "3." Otherwise, use the output of the second binary classifier DF_32v1 to determine the assignment category. If subject X is classified as "32" by binary classifier DF_32v1, assign the subject to category "2," otherwise assign it to category "1."

This sequence not only provides a unique category assignment for each subject, but the specific order used incorporates a stratification of risk into the classification process by identifying those subjects with a structural injury (and in need of the most urgent medical care) in the first discriminant step.

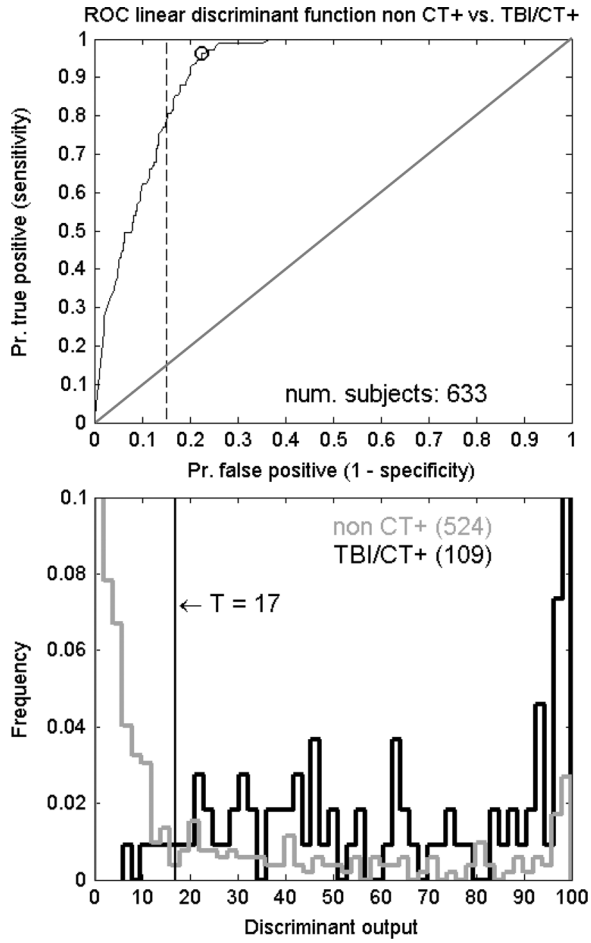


Fig. 4. Top panel: ROC curve for QDF discriminant function Category 3v21. Operating point selected for a target sensitivity of 95%. Bottom panel: Histogram of discriminant outputs/scores, with threshold indicated by "T," as determined in the top panel.

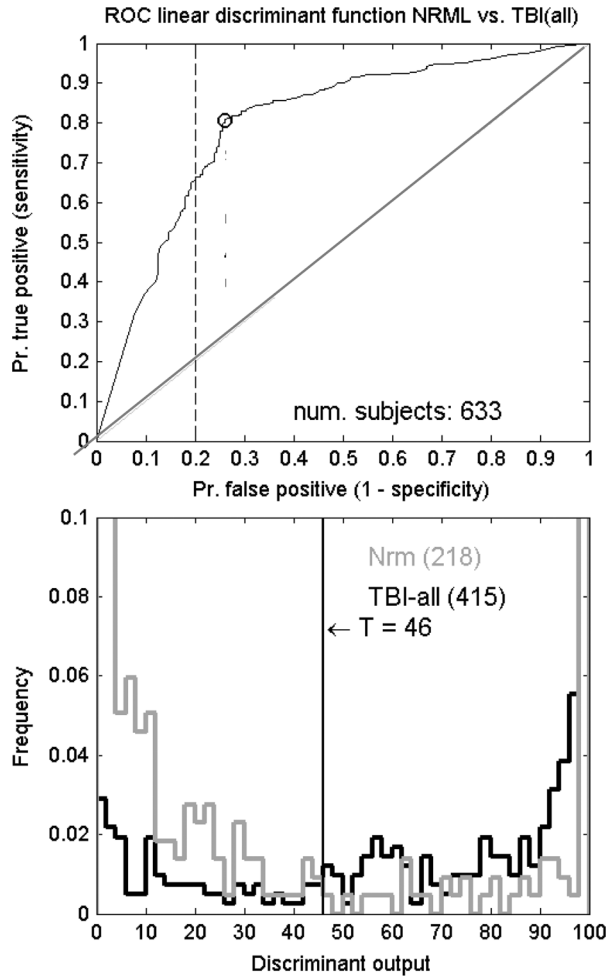


Fig. 5. Top panel: ROC curve for QDF discriminant function Category 32v1. Operating point selected for a target sensitivity of 95%. Bottom panel: Histogram of discriminant outputs/scores, with threshold indicated by "T," as determined in the top panel.

## V. Results

### *A. Individual Classifier Performance*

Following the prescribed procedure described above, the best performance for the two binary classification tasks was achieved by creating two QDF, one for classification of category 3v21 and a second for the classification of category 32v1. Target LOO sensitivities of 95% and 80%, respectively, were selected in creating the discriminants to reflect the desire for very high sensitivity in identifying the subjects in category 3 (structural injury visible on CT scan), and a balance between sensitivity and specificity for identification of the category 2 and 1 subjects. The QDF for 3v21 uses eighteen features (as determined by the method described above in Section IV-C "Rules to prevent over-fitting").

The variables that contributed most to this discrimination included monopolar and bipolar scale-free features for the total spectrum, absolute, and relative power features, especially in the theta frequency band, coherence features especially in high frequencies (alpha2, gamma), and fractal features. The QDF for 32v1 uses 28 features. Half of the variables that contributed most to this discrimination included measures of disturbances between regions including coherence, power asymmetry and phase synchrony, especially in the alpha band and total spectrum. Also contributing to this discrimination were monopolar and bipolar relative power across the spectrum, scale-free features for the total spectrum, and bipolar mean frequency especially in alpha and total spectrum.

Figs. 4 and 5 show the ROC curves for each discriminant and their corresponding histograms of discriminant outputs ($g(Z)$) calculated using the leave one out (LOO) framework. The LOO performance of each discriminant is shown in Tables II-A and II-B, including: sensitivity, specificity, PPV, NPV, AUC, and Cohen's d for the selected threshold values. High sensitivity (90%) with specificity of 80%, and NPV of 99%, was achieved.

In contrast, the performance of the QDF for 32v1, is more modest ($\mathrm{AUC} \simeq 0.80$). This reflects the clinical desirability for a balance between sensitivity and specificity, as well as the reality of greater overlap between the distributions of the categories being separated in this binary classification task (and hence the difficulty of separating them correctly). An illustration of the overlap between categories is shown in Fig. 6 for features SFAF7S and CoF1F2A2, in the 32v1 QDF, showing

TABLE II-A
PERFORMANCE OF CLASSIFIERS MEASURED USING SENSITIVITY AND SPECIFICITY IN TEN-FOLD AND LOO FRAMEWORK

| CLASSIFIER | 10-FOLD PERFORMANCE (SENSITIVITY, SPECIFICITY) | LOO PERFORMANCE (SENSITIVITY, SPECIFICITY) |
|---|---|---|
| 3 vs. 2,1 | 95.4%, 78.6% | 96.3%, 77.5% |
| 3,2 vs. 1 | 80.2%, 71.1% | 80.5%, 73.9% |

TABLE II-B
LOO PERFORMANCE OF CLASSIFIERS MEASURED USING SENSITIVITY SPECIFICITY, PPV, NPV, AUC, AND COHEN'S D

| | LOO PERFORMANCE | | |
|---|---|---|---|
| CLASSIFIER | SENSITIVITY, SPECIFICITY | PPV, NPV | AUC, COHEN'S D |
| 3 vs. 2,1 | 96.3%, 77.5% | 47.1%, 99.0% | 0.911, 1.94 |
| 3,2 vs. 1 | 80.5%, 73.9% | 85.4%, 66.5% | 0.797, 1.29 |

the distributions of z-scores of these features for all subjects in category 1 and category 2+3. The significant overlap between the distributions is evident. Such overlap supports the use of a quadratic classifier function since it enables separation based on both mean and standard deviation, whereas the performance of a linear classifier is dictated by separation in the means alone.

As an additional check of solution stability, as well as a demonstration of the effectiveness of the measures to prevent over-fitting, we compare the 10-fold and LOO performance results. Table II-A clearly demonstrates that 10-fold and LOO performance are very close, confirming good stability of the classifiers.

Since the quantitative features present in each classifier are age-regressed in order to remove the (normal) influence of age on these features, we would expect that the discriminant outputs and binary classification results are similarly not influence by age. In order to verify this, we split the subjects of the training database into four age bands: 15–25, 25–40, 40–55, 55 and over and plot the misclassification rates in each age band for the two binary QDF classifiers 3v21 and 32v1. Fig. 7 shows that the misclassification rates are nearly constant across the entire age range, confirming the effectiveness of our age regression methodology.

### B. Combined Classifier Performance

Beyond the output of the individual classifiers, we use the combined outputs of the two binary classifiers in conjunction with the category assignment logic described above to achieve three category classifications. These results can be summarized using a confusion matrix (see Tables III-A and III-B). In this matrix, each row represents the true category for each subject (truth) and each column represents the category determined by the discriminant algorithms (test). Hence, the main diagonal shows those subjects for whom the true and test categories agree and off-diagonal elements show misclassifications. Table III-B

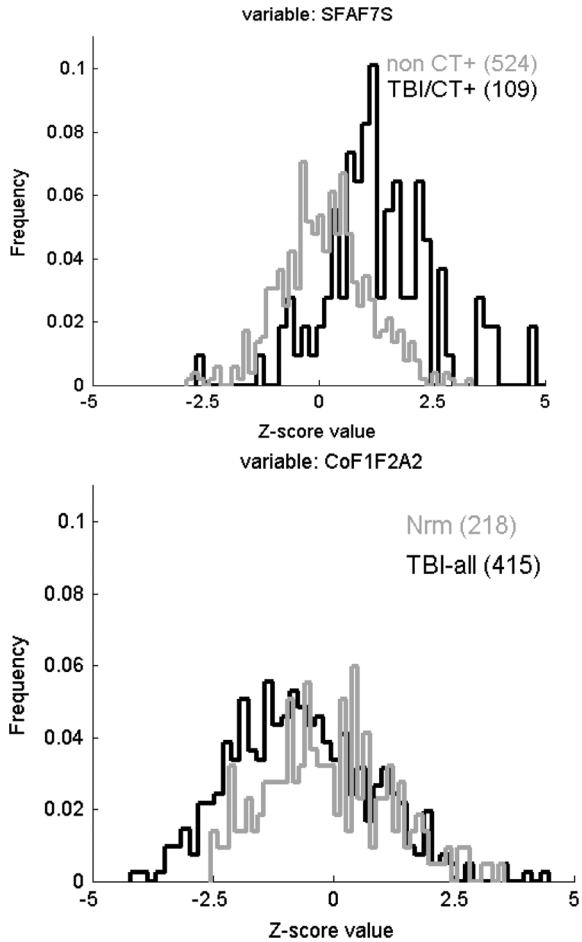


Fig. 6. Top panel: Distributions of z-transformed scale-free univariate variable SFAF7S for Normals (category 1) and for TBI subjects (category 2+3). Bottom panel: Distributions of z-transformed coherence univariate variable CoF1F2A2 for Normals (category 1) and for TBI subjects (category 2+3).

normalizes each entry by the total number of subjects in that category and expresses them as percentages.

## VI. DISCUSSION

Classification of mTBI patients into three clinically useful categories has been demonstrated using the proposed methods for discriminant classifier development. The classification of the CT+ group showed the expected high discrimination accuracy, as this group with structural brain injury on CT scan was most different from all the others (least overlapping). The target sensitivity was also set highest for this group, as the critical nature of the injury would not allow for many false negatives. The fact that specificity of 77% was achieved with this high sensitivity, with extremely high NPV showing that 99% of those called normal will in fact be normal upon scanning (absence of visible structural injury), further suggests that this could be a useful adjunct in the acute evaluation of head injuries and referrals for CT scans. In assessing the adequacy of this performance, the reported performance of other neuroimaging and screening tools should to be considered. The reported sensitivity for head CT in detecting acute stroke is approximately 78% [50]. Use of near-infrared spectroscopy (NIS) to assess the presence of a cerebral hematoma in patients presenting with mTBI, report

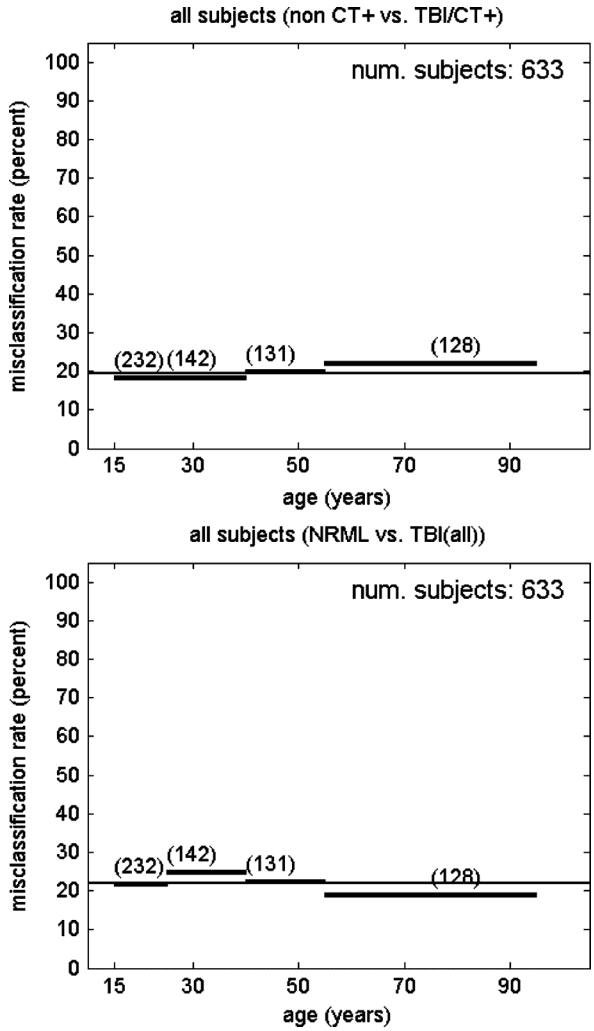


Fig. 7. Top panel: LOO misclassification rates for subjects in four age bands: 15–25, 25–40, 40–55, 55 and over for classifier QDF_3v21. Bottom panel: LOO misclassification rates for classifier QDF_32v1.

TABLE III-A
LOO CONFUSION MATRIX FOR CATEGORY ASSIGNMENT RULE (NUMBER OF SUBJECTS) BASED ON SEQUENCE OF 3 V 21 FOLLOWED BY 32 V 1

| TRUTH\TEST | 3 | 2 | 1 | N |
|---|---|---|---|---|
| 3 | **105*** | 2 | 2 | 109 |
| 2 | 85 | **157** | 64 | 306 |
| 1 | 33 | 38 | **147** | 218 |

* Numbers in bold type indicate numbers of subjects assigned to the correct category based on "Truth"

TABLE III-B
LOO CONFUSION MATRIX FOR CATEGORY ASSIGNMENT RULE (% OF SUBJECTS) BASED ON SEQUENCE OF 3 V 21 FOLLOWED BY 32 V 1

| TRUTH\TEST | CATEGORY 3 (%) | CATEGORY 2 (%) | CATEGORY 1 (%) |
|---|---|---|---|
| 3 | **96.3*** | 1.8 | 1.8 |
| 2 | 27.8 | **51.3** | 20.9 |
| 1 | 15.1 | 17.4 | **67.4** |

* Numbers in bold type indicate percentage (%) of subjects assigned to the correct category based on "Truth"

sensitivities and specificities around 80% [51]–[54], and performance was limited to specific volumes and distance from the brain surface. The sensitivity/specificity for the NOC and Canadian Head CT Rule (CCHR), clinical decision rules for obtaining a head CT for a suspected brain injury, are 100%/12.7% and 100%/50.6%, respectively, for detection of a clinically important brain injury, and achieve this high sensitivity at the expense of specificity [55]. The AUC of our discriminant was 0.91, demonstrating the ability of the prescribed process to create a discriminant function that enables accurate separation of the categories.

In the case of the identification of concussion, the sensitivity and specificity were lower, as expected due to the greater overlap between the categories, but still reached the sensitivity target of 80% with a specificity of 74%. From a neurophysiological or clinical perspective, the overlap between the categories and performance of the QDF_32v1 classifier also reflects the overlap the heterogeneity of milder traumatic brain injuries and less clear distinction between normal controls and mild concussion. Despite this, our performance is considerably higher than existing mTBI screening methods (primarily neurocognitive assessments) and the PPV of 85% further suggests that this method leads to a clinically useful evaluation tool that performs with high accuracy in the case of the presence of concussive injury. For example, the Military Acute Concussion Evaluation (MACE), a neurocognitive screening tool for mTBI, has a sensitivity of only 26% with a specificity of 88% [56], indicating that is not especially helpful in identifying injured soldiers, but rather only identifying those without injury.

The importance of the optimization of two different functions in the sequential application of classifiers allows for stratification-of-risk to be addressed separately in the discrimination of the three groups. That is, by first addressing the extreme risk of false negatives for the CT+ discriminant, together with the fact that this group has the least overlap with the others, CT+ could be identified with the highest targets for sensitivity and specificity. In the second function, where there is a significant overlap between the concussed patients and normal controls, and the "cost" of false positives is lower, appropriate targets can be met. A single function could not be optimized in this way.

The use of age regression of all QEEG features has been demonstrated previously in the scientific literature to greatly enhance the clinical utility of electroencephalography in the field of neuropsychiatry [57]. Among the important aspects of this approach is the fact that all features are converted to the common metric of probability and therefore can be combined into multivariate descriptors which can be used to describe brain "processes." The demonstration herein, that all classification was independent of age, supports the ability to use a single set of discriminant functions for all subjects as long as their age is known.

The informed data reduction step directly confronts a well-known problem for classification algorithms applied to conventional neuroimaging data (e.g., PET, MRI), that is, the small number of subjects compared with the enormous quantity of extracted features. Some researchers have ignored this problem completely, using thousands of variables as input to the classification algorithm, resulting in high classification accuracy with poor replicability or limited prospective validation. Others have used methods reliant on one dimension of feature reduction. In this paper we use a novel multidimensional approach including examination of the potential variables in terms of: replicability, separability, homogeneity of the distribution of each variable within the populations of interest, and assurance that the z-value for the feature in the more normal group is closer to zero. The resulting reduced feature pool then serves as input to a GA, a method uniquely suited to search problems in the biological domain, and not previously used in the literature in QEEG feature selection. In parallel, SFP, another feature selection method was implemented, which uses a deterministic approach to the selection problem. We note that although these two data mining methods were used in this work, there are other such methods not explored which could contribute to enhanced performance, and may be explored in future work. In recognition of the importance of the clinical applicability of the derived algorithms in solving the problem posed in this study, additional care was taken to reduce over-fitting and maximize use of the total subject population, and included cross-validation methods embedded in the algorithm development. Finally, selected features are input to two binary discriminant algorithms which are used to divide the population into the desired three categories.

The two discriminants both used a multidimensional combination of measure sets (absolute and relative power, power asymmetry, mean frequency, coherence, scale-free activity, phase synchrony, and fractal dimension), suggesting that the brain injury seen in this population would not have been well characterized by the use of one specific measure set, such as power in the lower frequency bands (e.g., delta and theta). Interestingly, while there is overlap in the features selected in the two functions, the selected feature sets are distinctive to the types of injury being discriminated. For example, half of the features selected in the classification of concussion contained measures of disturbances between frontal regions of the brain, including incoherence, power asymmetry and decreased phase synchrony (only present in this discriminant), emphasizing the role of functional connectivity and disturbances in power gradients in the pathophysiology of concussion. Since phase synchrony, has been demonstrated in the literature to correlate with injury to white matter tracts in the brain, a proposed etiological factor in concussion, it is confirmational that these features are selected in the classification of subjects with a concussion, Furthermore, in the work of Sponheim *et al.* (2011) [17], DTI findings correlated highest with frontal and frontotemporal phase synchrony of the EEG, supporting the adequacy of the limited montage used in this work. On the other hand, the classification of the structural injury group $(\mathrm{CT}+)$ contained both monopolar and bipolar scale-free features and absolute and relative power in the theta band, as would be expected from the literature in the description of slow wave features related to gray matter abnormalities [19]. It is also of note that Fractal dimension has been reported as a discriminating EEG feature in the classification in abnormal neuropsychiatric populations [58].

This proposed methodology resulted in algorithms that can be embedded in a portable device, using a limited montage of brain electrical signals, enabling clinical utility for triage of TBI in the acute setting. Implementing informed data reduction, and a sequence of algorithms that perform optimally for the relative risks in different segments of the patient population, resulted in high discriminant accuracy. Prospective validation in a broader multicenter trial is currently underway and could provide the necessary evidence of clinical validity to introduce such methods into the clinical community.

## APPENDIX

### A. Derivation of Nontraditional Univariate Features

1) *Fractal dimension:* The fractal dimension of a 1-D signal is the fractal dimension of the curve formed by the plot of that signal. It is a nonlinear mathematical quantity (taking values between 1 and 2), which reflects to what degree the curve fills 2-D space as one zooms down to finer and finer scales. The feature is computed as the average fractal dimension (FD) of the temporal EEG signal, obtained from successive FD estimates computed over segments of length 256 samples. These estimates are computed using the method proposed by Higuchi [59], with the parameter $k\_\max$ set to 6.
2) *2. Scale-free brain activity:* The power spectrum of spontaneous brain activity typically follows a $1/\mathrm{f}^{\beta}$ law over the frequency range 1–25 Hz and is therefore approximately linear when plotted on a log-log scale. This behavior of the power spectrum is typical of fractal processes. The parameter $\beta$ is usually referred to as *scale-free brain activity* [38]. We compute this parameter by fitting a line to the graph of $\log(P(f))$ versus $\log(f)$, where $f$ denotes the frequency in hertz and $P(f)$ denotes the estimate of the power spectral density of the EEG at frequency $f$, computed using artifact-free epochs.
3) *Information theory-based measures:* Our database includes four types of these measures: traditional Shannon entropy, Tsallis entropy, wavelet entropy and relative wavelet entropy. The traditional Shannon entropy of time series $x[n]$ is computed by first partitioning the amplitude range of the EEG into $M$ bins $I_i$ of width 5 $\mu$V. The probability distribution $p_i$ is obtained by the calculating the ratio of the number of data samples falling into each bin $I_i$. The formula for the Shannon entropy of the signal is

$$ES(x) = -\sum_{i=1}^{M} p_i \ln(p_i) \tag{A1}$$

After the same binning procedure as described above, the Tsallis entropy (computed for parameter $q = 3$ and

$q = 5$) of time series $x[n]$ is computed from the following formula:

$$TE\_q(x) = \left(1 - \sum_{i=1}^{M} p_i^q\right) / (q-1) \tag{A2}$$

The computation of the (total) wavelet entropy of time series $x[n]$ first requires the computation of the complex wavelet transform of $x[n]$. For this we used the complex wavelet filters proposed by Selesnick (2002) [60]. We denote by $C_j(k)$ the complex coefficient of the wavelet transform at resolution level (scale) $j$. Following the definitions given in Rosso *et al.*, (2001) [61], the energy of the detail signal at scale $j$ is denoted by $E_j$ and the total energy by $E$. They are computed according to

$$E_j(x) = \sum_k |C_j(k)|^2, \quad E = \sum_j E_j \tag{A3a}$$

The normalized values of $E_j$ represent the *relative wavelet energy* denoted by $p_j$.

$$p_j = E_j / E \tag{A3b}$$

The distribution $p_j$ can therefore be considered a time-scale density for which the (Shannon) entropy is defined as

$$\mathrm{WET}(x) = -\sum_j p_j \ln(p_j) \tag{A3c}$$

This measure evaluates the complexity of the energy distribution across frequency bands [33].

The computation of the (relative) wavelet entropy of time series $x[n]$ and $y[n]$ similarly first requires the computation of the complex wavelet transforms of $x[n]$ and $y[n]$. Let $p_j$ and $q_j$ denote the respective distributions of wavelet energies of $x$ and $y$. The relative wavelet entropy is defined as [61]

$$\mathrm{WRA}(x, y) = \sum_j p_j \ln(p_j / q_j). \tag{A4}$$

4) *Functional connectivity measures:* Phase delay (or lag) was shown by Thatcher *et al.* to be a useful feature for the task of separating normal subjects from those with mild head trauma [15]. Our computation of phase delay is performed as described in the work of Thatcher *et al.* [62].

The Phase Synchrony measure described by Aviyente *et al.* [63] proposes to measure the neural coordination in the brain. In this sense it can be seen as a measure of *brain connectivity*. This measure first requires the computation of a time-frequency distribution (TFD) such as the RID-Rihaczek distribution for a pair of signals $x[n]$ and $y[n]$. We denote by $C_1(n, m), C_2(n, m)$ the TFDs of $x[n]$ and $y[n]$, respectively. Then, the time-varying phase spectrum of $x$ and $y$ is computed as follows [see [63] (10)]:

$$\phi_{12}(n, m) = \arg[C_1(n, m) C_2^*(n, m) / |C_1(n, m)||C_2(n, m)|] \tag{A5a}$$

The Phase Locking Value across $N$ epochs of EEG signals $x$ and $y$ is then defined as the inter-epoch variability of this phase spectrum [see [63] (12)], namely

$$\mathrm{PLV}(n, m) = 1/N \left| \sum_{k=1}^{N} \exp(j \phi_{12}^k(n, m)) \right| \tag{A5b}$$

The phase synchrony for a given frequency band is finally defined as the average of the phase locking value (PLV) over that frequency band.

### B. Figures of Merit and Receiver Operating Characteristic

1) *Definition of FOM: Sensitivity, Specificity, PPV, and NPV:* In the context of *medical diagnostic* where the goal is to determine with accuracy the presence or absence of a specific disease or condition, it is useful to differentiate between "state A" versus "state B," where states A and B can, for example, represent respectively: "disease present" and "disease absent." For each subject in a sample group where true classification information (diagnosis) is available, we can determine whether the output of the classifier function (also called "output of the test") matches the diagnosis or not. This leads, for each subject, to incrementing one counter among the following four: "true positives" (subjects with the disease for which the test is positive), "true negatives" (subjects without the disease for which the test is negative), "false positives" (subjects without the disease for which the test is positive), and "false negatives" (subjects with the disease for which the test is negative).
*Sensitivity* of the classifier (or of the test) is the ratio of "true positives" over the number of subjects for whom "disease" is present. *Specificity* of the test is the ratio of "true negatives" over the number of subjects for whom disease is absent. *PPV* is the probability that disease is present when the test result is positive ("true positives" over the number of subjects for whom the test result is positive). *NPV* is the probability that disease is absent when the test result is negative ("true negatives" over the number of subjects for whom the test result was negative.) In the context of our two binary classifiers, 3 v 21, 32 v 1, we adopt the convention that the "most serious brain injury condition" corresponds to the disease condition, (state A). Therefore, for 3 v 21, the disease condition is "3," for 32 v 1, the disease condition is "32." All the sensitivity and specificity numbers reported in this paper should be understood with this convention in mind.
2) *Discriminant Output and ROC Curves:* The output of each of our discriminant functions is a number which can take any value between 0 and 100. Once a *critical value* (or threshold) T is selected, the output of the test becomes binary and sensitivity and specificity for this particular threshold can be computed. The ROC is the curve through the set of points: $\{(1 - \mathrm{specificity(T)}, \mathrm{sensitivity(T)})\}$, which is obtained by varying the value of the critical value T between 0 and 100.

The ROC curve is therefore a graphical illustration of the achievable statistical performance of a given test/discriminant, depending on the selected critical value. For any discriminant

described in this document, we show ROC curves and histograms of the discriminant out.

## ACKNOWLEDGMENT

The authors would like to acknowledge the dedicated hard work of all the clinical sites and clinical coordinators for acquisition of the data used in this work.

## REFERENCES

[1] P. R. Harper, "A review and comparison of classification algorithms for medical decision making," *Health Policy*, vol. 71, no. 3, pp. 315–331, Mar. 2005.

[2] D. Lehmann, P. L. Faber, S. Galderisi, W. M. Herrmann, T. Kinoshita, M. Koukkou, A. Mucci, R. Pascual-Marqui, N. Saito, J. Wackermann, G. Winterer, and P. Koenig, "EEG microstate duration and syntax in acute, medication-naive, first-episode schizophrenia: A multi-center study," *Psychiatry Res.*, vol. 138, no. 2, pp. 141–156, 2005.

[3] R. L. Somorjai, B. Dolenko, A. Nikulin, W. Roberson, and N. Thiessen, "Class proximity measures—Dissimilarity-based classification and display of high-dimensional data," *J. Biomed. Informat.*, vol. 44, no. 5, pp. 775–788, Oct. 2011.

[4] L. S. Prichep, E. R. John, S. H. Ferris, L. Rausch, Z. Fang, L. Rausch, R. Cancro, C. Torrossian, and B. Reisberg, "Prediction of longitudinal cognitive decline in normal elderly with subjective complaints using electrophysiological imaging," *Neuro. Biol. Aging*, vol. 27, pp. 471–481, 2006.

[5] S. C. Suffin, W. H. Emory, N. Gutierrez, G. S. Arora, M. J. Schiller, and A. Kling, "A QEEG database method for predicting pharmacotherapeutic outcome in refractory major depressive disorder," *J. Am. Physicians Surgeons*, vol. 12, no. 4, pp. 104–108, 2007.

[6] R. W. Thatcher, D. M. North, R. T. Curtin, R. A. Walker, C. J. Biver, J. F. Gomez, and A. M. Salazar, "An EEG severity index of traumatic brain injury," *J. Neuropsychiatry Clin. Neurosci.*, vol. 13, no. 1, pp. 77–81, 2001.

[7] M. Rabinoff, C. M. R. Kitchen, I. A. Cook, and A. F. Leuchter, "Evaluation of quantitative EEG by classification and regression trees to characterize responders to antidepressant and placebo treatment," *Open Med. Inform. J.*, vol. 5, pp. 1–8, 2011.

[8] M. Faul, L. Xu, M. M. Wald, and V. G. Coronado, Traumatic brain injury in the United States: Emergency department visits, hospitalizations, and deaths 2002–2006. Atlanta, GA, 2010.

[9] J. A. Langlois, W. Rutland-Brown, and M. M. Wald, "The epidemiology and impact of traumatic brain injury: A brief overview," *J. Head Trauma Rehabil.*, vol. 21, no. 5, 2006.

[10] D. L. Trudeau, J. Anderson, L. Hansen, B. A. Shagalov, M. M. E. Schmoller, S. Nugent, and S. Barton, "Findings of mild traumatic brain injury in combat veterans with PTSD and a history of blast concussion," *J. Neuropsychiatry Clin. Neurosci.*, vol. 10, pp. 308–313, 1998.

[11] I. G. Stiell, G. A. Wells, K. Vandemheen, C. Clement, H. Lesiuk, A. Laupacis, R. D. McKnight, R. Verbeek, R. Brison, D. Cass, M. A. Eisenhauer, G. H. Greenberg, and J. Worthington, "The Canadian CT head rule for patients with minor head injury," *Lancet*, vol. 357, no. 9266, pp. 1391–1396, May. 2001.

[12] S. C. Stein, A. Fabbri, F. Servadei, and H. A. Glick, "A critical comparison of clinical decision instruments for computed tomographic scanning in mild closed traumatic brain injury in adolescents and adults," *Ann. Emerg. Med.*, vol. 53, no. 2, pp. 180–188, Feb. 2009.

[13] J. R. Wozniak, L. Krach, E. Ward, B. A. Mueller, R. Muetzel, S. Schnoebelen, A. Kiragu, and K. O. Lim, "Neurocognitive and neuroimaging correlates of pediatric traumatic brain injury: A diffusion tensor imaging (DTI) study," *Arch. Clin. Neuropsychol.*, vol. 22, no. 5, pp. 555–568, Jun. 2007.

[14] R. W. Thatcher, C. Biver, R. Mc Alaster, M. Camacho, and A. Salazar, "Biophysical linkage betweeen MRI and EEG amplitude in closed head I," *Neuroimage*, vol. 7, pp. 352–367, 1998.

[15] R. W. Thatcher, R. A. Walker, I. Gerson, and F. H. Geisler, "EEG discriminant analyses of mild head trauma," *EEG Clin. Neurophysiol.*, vol. 73, pp. 94–106, 1989.

[16] K. H. Taber, D. L. Warden, and R. A. Hurley, "Blast-related traumatic brain injury: What is known?," *J. Neuropsychiatry Clin. Neurosci.*, vol. 18, no. 2, pp. 141–145, 2006.

[17] A. R. Sponheim, K. A. McGuire, S. S. Kang, N. D. Davenport, S. Aviyente, E. M. Bernat, and K. L. Lim, "Evidence of disrupted functional connectivity in the brain after combat-related blast injury," *Neuroimage*, vol. 54, pp. s21–s29, 2011.

[18] C. L. Mac Donald, M. Johnson, D. Cooper, E. C. Nelson, N. J. Werner, J. S. Shimony, A. Z. Snyder, M. E. Raichle, J. R. Witherow, R. Fang, S. F. Flaherty, and D. L. Brody, "Detection of blast-related traumatic brain injury in U.S. military personnel," *N. Eng. J. Med.*, vol. 364, no. 22, pp. 2091–2100, Jun. 2011.

[19] C. Cao, R. I. Tutwiler, and S. Slobounov, "Automatic classification of athletes with residual functional deficits following concussion by means of EEG signal using support vector machine," *Neural Syst. Rehabil. Eng.*, vol. 16, no. 4, pp. 327–335, 2008.

[20] H. Levin, D. Williams, H. Eisenberg, W. M. High, and F. C. Guinto, "Serial MRI and neurobehavioural findings after mild to moderate closed head injury," *J. Neurol., Neurosurg. Psychiatry*, vol. 55, pp. 255–262, 1992.

[21] P. F. Malloy and M. Aloia, "Frontal lobe dysfunction in traumatic brain injury," *Seminars Clini. Neuropsychiatry*, vol. 3, no. 3, pp. 186–194, 1998.

[22] O. Duda, P. E. Hart, and D. G. Stok, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[23] D. M. J. Tax and R. P. W. Duin, "Using two-class classifiers for multiclass classification," in *Proc. 2nd Int. Conf. Pattern Recognit.*, 2002, pp. 124–127.

[24] G. Teasdale and B. Jennett, "Assessment of coma and impaired consciousness," *Lancet*, vol. 2, pp. 81–83, 1974.

[25] R. C. Cantu, "Posttraumatic retrograde and anterograde amnesia: Pathophysiology and implications in grading and safe return to play," *J. Athletic Train.*, vol. 36, no. 3, pp. 244–248, 2001.

[26] D. Erlanger, T. Kaushik, R. Cantu, J. T. Barth, D. K. Broshek, J. R. Freeman, and F. M. Webbe, "Symptom-based assessment of the severity of a concussion," *J. Neurosurg.*, vol. 98, no. 3, pp. 477–484, Mar. 2003.

[27] E. M. Whitham, K. J. Pope, S. P. Fitzgibbon, T. Lewis, C. R. Clark, S. Loveless, M. Broberg, A. Wallace, D. DeLosAngeles, P. Lillie, A. Hardy, R. Fronsko, A. Pulbrook, and J. O. Willoughby, "Scalp electrical recording during paralysis: Quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG," *Clin. Neurophysiol.*, vol. 118, no. 8, pp. 1877–1888, Aug. 2007.

[28] A. Jacquin, E. Causevic, and E. R. John, "Automatic identification of spike wave events and non convulsive seizures with a reduced set of electrodes," in *29th Annu. Int. Conf. IEEE EMBS*, 2007, pp. 1928–1931.

[29] B. J. Fisch, *Fisch and Spehlmann's EEG Primer*, 3rd ed. Amsterdam: Elsevier, 1999, pp. 3–17.

[30] R. Isenhart, A. Jacquin, B. Howard, J. Filipenko, and L. S. Prichep, "Automatic detection of artifacts in frontal brain EEG signals," *Comp. Med. Biol.*, 2012, submitted for publication.

[31] R. G. Norman, I. Pal, C. Stewart, J. A. Walsleben, and D. M. Rapoport, "Interobserver agreement among sleep scorers from different centers in a large dataset," *Sleep*, vol. 23, no. 7, pp. 901–908, 2000.

[32] E. R. John, L. S. Prichep, and P. Easton, , A. S. Gevins and A. Remond, Eds., "Normative data banks and neurometrics: Basic concepts, methods and results of norm construction," in *Handbook of Electroencephalography and Clinical Neurophysiology*. Amsterdam, The Netherlands: Elsevier, 1987, vol. I, pp. 449–495.

[33] N. V. Thakor and S. Tong, "Advances in quantitative electroencephalogram analysis methods," *Annu. Rev. Biomed. Eng.*, vol. 6, no. 1, pp. 453–495, Aug. 2004.

[34] V. Sakkalis, "Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG," *Comput. Biol. Med.*, vol. 41, no. 12, pp. 1110–1117, Dec. 2011.

[35] E. R. John, H. Ahn, L. S. Prichep, M. Trepetin, D. Brown, and H. Kaye, "Developmental equations for the electroencephalogram," *Science*, vol. 210, pp. 1255–1258, 1980.

[36] E. R. John, L. S. Prichep, J. Friedman, and P. Easton, "Neurometrics: Computer-assisted differential diagnosis of brain dysfunctions," *Science*, vol. 293, pp. 162–169, 1988.

[37] T. Gasser, P. Bacher, and J. Mochs, "Transformation towards the normal distribution of broad band spectral parameters of the EEG," *EEG Clin. Neurophysiol.*, vol. 53, no. 1, pp. 119–124, 1982.

[38] B. J. He, J. M. Zempel, A. Z. Snyder, and M. E. Raichle, "The temporal structures and functional significance of scale-free brain activity," *Neuron*, vol. 66, no. 3, pp. 353–369, May 2010.

[39] R. W. Thatcher, D. North, and C. Biver, "EEG and intelligence: Relations between EEG coherence, EEG phase delay and power," *Clin. Neurophysiol.*, vol. 116, no. 9, pp. 2129–2141, Sep. 2005.
[40] J. M. Weiner and O. J. Dunn, "Elimination of variates in discrimination problems," *Biometrics*, vol. 22, pp. 268–275, 1966.
[41] A. Kondacs and M. Szabo, "Long-term intra-individual variability of the background EEG in normals," *Clin. Neurophysiol.*, vol. 110, p. 1708, 1999.
[42] M. Mitchell, *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press, 1996.
[43] W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection," *Pattern Recognit. Lett.*, vol. 10, no. 5, pp. 335–347, Nov. 1989.
[44] J. H. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," *IEEE Intell. Syst.*, vol. 13, no. 2, pp. 44–49, Mar./Apr. 1998.
[45] M. L. Raymer, E. D. Punch, E. D. Goodman, L. A. Kuhn, and A. K. Jain, "Dimensionality reduction using genetic algorithms," *IEEE Trans. Evolut. Computat.*, vol. 4, no. 2, pp. 164–171, Jul. 2000.
[46] I. S. Oh, J. S. Lee, and B. R. Moon, "Hybrid genetic algorithms for feature selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1424–1437, Nov. 2004.
[47] J. Li and J. P. Fine, "ROC analysis with multiple classes and multiple tests: Methodology and its application in microarray studies," *Biostatistics*, vol. 9, no. 3, pp. 566–576, July 2008.
[48] S. J. Raudys and A. K. Jain, "Small sample size effects in statistical pattern recognition: Recommendations for practitioners," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 3, pp. 252–264, Mar. 1991.
[49] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Int. Joint Conf. Artif. Intell.*, 1995, pp. 1137–1145.
[50] M. A. Kalafut, D. L. Schriger, J. L. Saver, and S. Starkman, "Detection of early CT signs of $>1/3$ middle cerebral artery infarctions: Interrater reliability and sensitivity of CT interpretation by physicians involved in acute stroke care," *Stroke*, vol. 31, no. 7, pp. 1667–1671, 2000.
[51] S. Kahraman, H. Kayali, C. Atabey, F. Acar, and S. Gocmen, "The accuracy of near-infrared spectroscopy in detection of subdural and epidural hematomas," *J. Trauma Acute Care Surg.*, vol. 61, no. 6, 2006.
[52] B. Kessel, I. Jeroukhimov, I. Ashkenazi, T. Khashan, M. Oren, J. Haspel, M. Medvedev, V. Nesterenko, A. Halevy, and R. Alfici, "Early detection of life-threatening intracranial haemorrhage using a portable near-infrared spectroscopy device," *Injury*, vol. 38, no. 9, pp. 1065–1068, Sept. 2007.
[53] H. Ghalenoui, H. Saidi, M. Azar, S. T. Yahyavi, H. B. Razavi, and M. Khalatbari, "Near-infrared laser spectroscopy as a screening tool for detecting hematoma in patients with head trauma," *Prehospital Disaster Med.*, vol. 23, no. 6, pp. 558–561, 2008.
[54] J. Leon-Carrion, J. M. Dominguez-Roldan, U. Leon-Dominguez, and F. Murillo-Cabezas, "The infrascanner, a handheld device for screening in situ for the presence of brain haematomas," *Brain Inj.*, vol. 24, no. 10, pp. 1193–1201, Aug. 2010.
[55] 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, and G. A. Wells, "Comparison of the Canadian CT head rule and the New Orleans criteria in patients with minor head injury," *JAMA*, vol. 294, no. 12, pp. 1511–1518, Sep. 2005.
[56] R. L. Coldren, M. P. Kelly, R. V. Parish, M. Dretsch, and M. L. Russell, "Evaluation of the military acute concussion evaluation for use in combat operations more than 12 hours after injury," *Military Med.*, vol. 175, no. 7, p. 477, 2010.
[57] L. S. Prichep, "Use of normative databases and statistical methods in demonstrating clinical utility of QEEG: Importance and cautions," *Clin. EEG*, vol. 36, no. 2, pp. 82–87, 2005.
[58] B. S. Raghavendra, D. N. Dutt, H. N. Halahalli, and J. P. John, "Complexity analysis of EEG in patients with schizophrenia using fractal dimension," *Physiolog. Measure.*, vol. 30, no. 8, pp. 795–808, 2005.
[59] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D*, vol. 31, pp. 277–283, 1988.
[60] I. W. Selesnick, A new complex-directional wavelet transform and Its application to image denoising 2002, pp. 573–576.
[61] O. A. Rosso, S. Blanco, J. Yordanova, V. Kolev, A. Figliola, M. Schurmann, and E. Basar, "Wavelet entropy: A new tool for analysis of short duration brain electrical signals," *J. Neurosci. Methods*, vol. 105, no. 1, pp. 65–75, Jan. 2001.
[62] R. W. Thatcher, D. North, and C. Biver, "Evaluation and validity of a LORETA normative EEG database," *Clin. EEG Neurosci.*, vol. 36, no. 2, pp. 116–122, 2005.
[63] S. Aviyente, E. M. Bernat, W. S. Evans, and S. R. Sponheim, "A phase synchrony measure for quantifying dynamic functional integration in the brain," *Human Brain Mapp.*, vol. 32, no. 1, pp. 80–93, 2011.

**Leslie S. Prichep** received the B.S. and M.S. degrees in biopsychology from the University of Pittsburgh, Pittsburgh, PA, in 1968 and 1969, respectively, and the Ph.D. degree in experimental psychobiology from City University of New York, New York, in 1974.

Her major field of study is translational research in quantitative electrophysiology. She is the Associate Director of Research Laboratories (BRL) and a Professor of Psychiatry at New York University School of Medicine, New York, and serves as a scientific consultant to BrainScope Biomedical Instrument Company. She has extensive research experience using quantitative analysis of brain electrical activity in the study of neuropsychiatric and cognitive disorders. She has over 100 publications and is considered one of the leading research scientists in the field of computerized electrophysiology.

**Arnaud E. Jacquin** (M'88–SM'05) received the Diplôme d'ingéniéur de l'Ecole Supérieure d'Electricité (B.S. degree in electrical engineering from SUPELEC), Gifsur-Yvette, France in 1986, and the M.Sc. Degree in electrical engineering and the Ph.D. degree in mathematics from the Georgia Institute of Technology, Atlanta, in 1988 and 1989, respectively.

He was a Senior DSP Engineer at Lucent Bell Laboratories, N.J., from 1990 to 2003. In 2003, he joined Everest Biomedical Instrument Company, where he was a Senior DSP Engineer. In 2006 he joined BrainScope Company Inc., where he is currently Director of Algorithm Development involved in the design of algorithms for portable medical devices using EEG. His research interests are in the field of signal processing including biosignal denoising and analysis, pattern analysis, classification, and optimization. He is a co-inventor of 14 U.S. patents and has published more than 35 refereed articles. He was awarded the IEEE Signal Processing Society Senior Award (Image and Multidimensional Signal Processing Area), in 1993.

**Julie Filipenko** received the B.S. degree in biomedical engineering from Rensselaer Polytechnic Institute, Troy, NY, in 2001, and the M.S. degree in biomedical engineering from Columbia University, New York, in 2005.

She joined Accenture, New York, in 2001, and was a Member of the High Technology and Communication operating group. In 2004, she joined Everest Biomedical Instrument Company where she was a Consultant. In June 2006, she joined BrainScope Company Inc., where she is currently Senior Clinical Database Manager and is involved in the design of electronic Case Report Forms and database management/data analysis. Her research interests are in the field of signal and image processing, biomedical imaging, mathematical analysis and quantification of medical images, classification, and optimization.

**Samanwoy Ghosh Dastidar** received the B.E. degree in civil engineering from the Indian Institute of Technology, Roorkee, India, in 1997, and the M.S. degree of science in civil engineering and the Ph.D. degree in biomedical engineering from The Ohio State University, Columbus, in 2002 and 2007, respectively.

His major field of study was biomedical signal and image processing and neuroscience. From 1999 to 2007, he conducted research and taught at The Ohio State University. In 2007, he joined ANSAR Med-

ical Technologies as Principal Biomedical Engineer, leading a cross-functional team for the design, development, and miniaturization of an autonomic nervous system monitoring technology. In 2010, he joined BrainScope Company Inc., as an Integration Engineer, focusing on the development of classification algorithms for the detection of TBI and translation of the algorithms into real time devices. He has over 20 publications and book chapters.

Dr. Ghosh Dastidar is a member of Sigma Xi and the Biomedical Engineering Society.

**Stephen Zabele** received the B.E. degree in electrical engineering, the M.S. degree in electrical engineering, and the Ph.D. degree in electrical engineering from Clarkson University, in 1977, 1979, and 1984, respectively. His major field of study was in communications and signal processing.

From 1979 to 1981, he worked on radar signal processing problems for Raytheon. From 1981 to 1984 he held a faculty at Clarkson University. In 1984, he joined TASC, Reading, MA. In 2001, he joined ALPHATECH, where he rose to the position of Chief Scientist. In 2008, he joined BBN Technologies as Vice President of Technology Applications, where he has supported technology development activities across the organization, ranging from healthcare technologies to middleware infrastructures for network applications. He holds several patents in diverse technology areas, has over 30 publications including several invited papers, and a formal IETF RFC.

**Asmir Vodenčarević** (M'06) received the Diplomirani inženjer elektrotehnike (Dipl.-Ing. degree in electrical engineering, with honors) from the University of Tuzla, Bosnia and Herzegovina, in 2007. He has been member of the International Graduate School Dynamic Intelligent Systems since April 2010, working toward the Ph.D. degree in a research group Knowledge-Based Systems, the University of Paderborn, Paderborn, Germany.

He joined Novagenus d.o.o., Tuzla, BiH, in October 2006, where he worked as the Internet Programmer. In July 2007, he joined Inspire Company d.o.o., Tuzla, BiH (subsidiary of BrainScope Co. Inc.), where he was the Embedded Software Developer. In June 2009, he joined Pointes d.o.o. Tuzla, BiH, where he continued software development for medical devices. His research interests are algorithm development, machine learning, classification, evolutionary computation, and artificial intelligence.

**Neil S. Rothman** received the B.S. degree in biomedical engineering and the M.S. degree in mechanical engineering from Rensselaer Polytechnic Institute, Troy, NY, in 1978 and 1978, respectively, and the Ph.D. degree in mechanical engineering from The Johns Hopkins University, Baltimore, MD, in 1994.

He was a consultant to GE Healthcare's Maternal and Infant Care Division, prior to joining Infinite Biomedical Technologies, where he was Vice President of R&D leading the development of EEG based systems. He was also the Vice President of Operations for Metasensors, Director of Engineering for IGEN International, and Senior Vice President and Chief Technical Officer for CardioLogic Systems. He also taught at the Johns Hopkins University, Baltimore, MD, in the areas of technology commercialization and engineering management. He joined BrainScope, Inc., in 2009, where he is the Vice President of Product Development. He also has 12 patents.