Evaluating Performance of EEG Data-Driven Machine Learning for Traumatic Brain Injury Classification

Michael Caiola[1], Nico Vivaldi[2], & Meijun Ye[1]

[1]FDA, CDRH, Office of Science and Engineering Laboratories, Division of Biomedical Physics. [2]Mathworks, Natick, MA.

Abstract

Purpose: We evaluated the performance of commonly used supervised machine learning algorithms in the classification of patients with traumatic brain injury history from those with stroke history and/or normal EEG. **Methodology:** Support vector machine (SVM) and K-nearest neighbors models were generated with a diverse feature set (selected using various feature section techniques) from Temple EEG Corpus for both two-class classification of patients with TBI history from normal subjects and three-class classification of TBI, stroke and normal subjects. **Results:** For two-class classification, an accuracy of 0.94 was achieved in 10-fold cross validation (CV), and 0.76 in independent validation (IV). For three-class classification, 0.85 and 0.71 accuracy were reached in CV and IV respectively. Overall, linear discriminant analysis (LDA) feature selection and SVM models consistently performed well in both CV and IV and for both two-class and three-class classification. Compared to normal control, both TBI and stroke patients showed an overall reduction in coherence and relative PSD in delta frequency, and an increase in higher power. But stroke patients showed a greater degree of change and had additional global decrease in theta power. **Conclusion:** Our study suggests that EEG data-driven machine learning can be a useful tool for TBI classification and can potentially provide specificity to separate different neurological conditions.

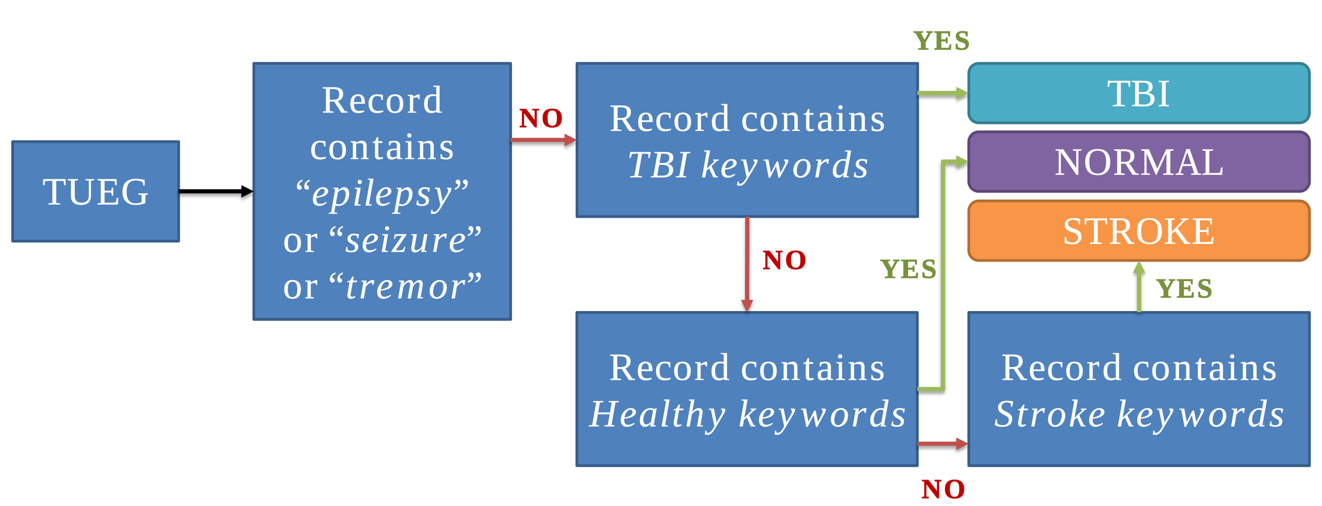
Introduction

Traumatic brain injury (TBI) presents a significant challenge affecting an estimated 2.5 million people annually. Current clinical scores can classify TBI by severity but not with enough sensitivity to detect mild TBI or monitor progression. Therefore, efforts are ongoing to seek for alternative clinical assessment tools for TBI.

EEG has advantages of being non-invasive, easy-to-use, portable and cost effective. However, when applied to TBI research, EEG yields mixed results in the literature with some studies showing significant differences in EEG-based power spectra data between mild TBI and normal groups, while others report no such distinction.

Due to the inherent complexity of TBI, including the absence of consensus on biomarkers, underlying relationships between data, and patient-to-patient variability, big data analytics have the potential to make determinations about population characteristics that would otherwise be too difficult/impossible to manually identify.

The Temple University Hospital's EEG Corpus (TUEG) is a rich archive of over 30,000 clinical EEG recordings collected at Temple University Hospital from 2002 to present. Using key word text matching, a cohort of TBI patients (N=142), a cohort of patients who had strokes (N=165), and a cohort of normal patients (N=105) were identified.



Materials and Methods

For each subject, 3 minutes of awake, resting-state, stimuli-free EEG recordings were included. Signals (fs = 250 Hz) from 19 common channels in ten-twenty standard arrangement (FP1, FP2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, FZ, CZ, PZ) were filtered with cutoff frequency passband 1-100 Hz. Artifact rejection was done using Independent Component Analysis (ICA). All analysis was done on the **raw** data and **clean** (ICA artifact-removed) data.

Phase-amplitude coupling (PAC), absolute and relative power spectral density (PSD) within frequency bands, spectral entropy (SE), and inter-channel cross coherence (Coh) were calculated for both raw and clean data, resulting in 1330 EEG features plus 3 demographic features.

For independent validation (IV), each dataset was split into a training and test set. The feature dimensionality was then reduced using the training dataset and a range of feature selection techniques: statistics, Linear Discriminant Analysis (LDA), Forward Sequential Feature Selection (FSFS), Backwards Sequential Feature Selection (BSFS), and Principal Component Analysis (PCA). These reduced feature sets fed into Support Vector Machine (SVM) models using six different kernels and K-Nearest Neighbors (KNN) classifiers using six definitions for training. Accuracy was determined by 10-fold cross validation (CV), IV, and randomly labeled data.

Two variants of these models were trained: Normal vs TBI (**2-class**) and Normal vs TBI vs Stroke (**3-class**).

Models were trained and tested on both raw and clean data, data excluding demographic information and five different selection methods for 2-class and 3-class classification for a total of 480 models.

Results

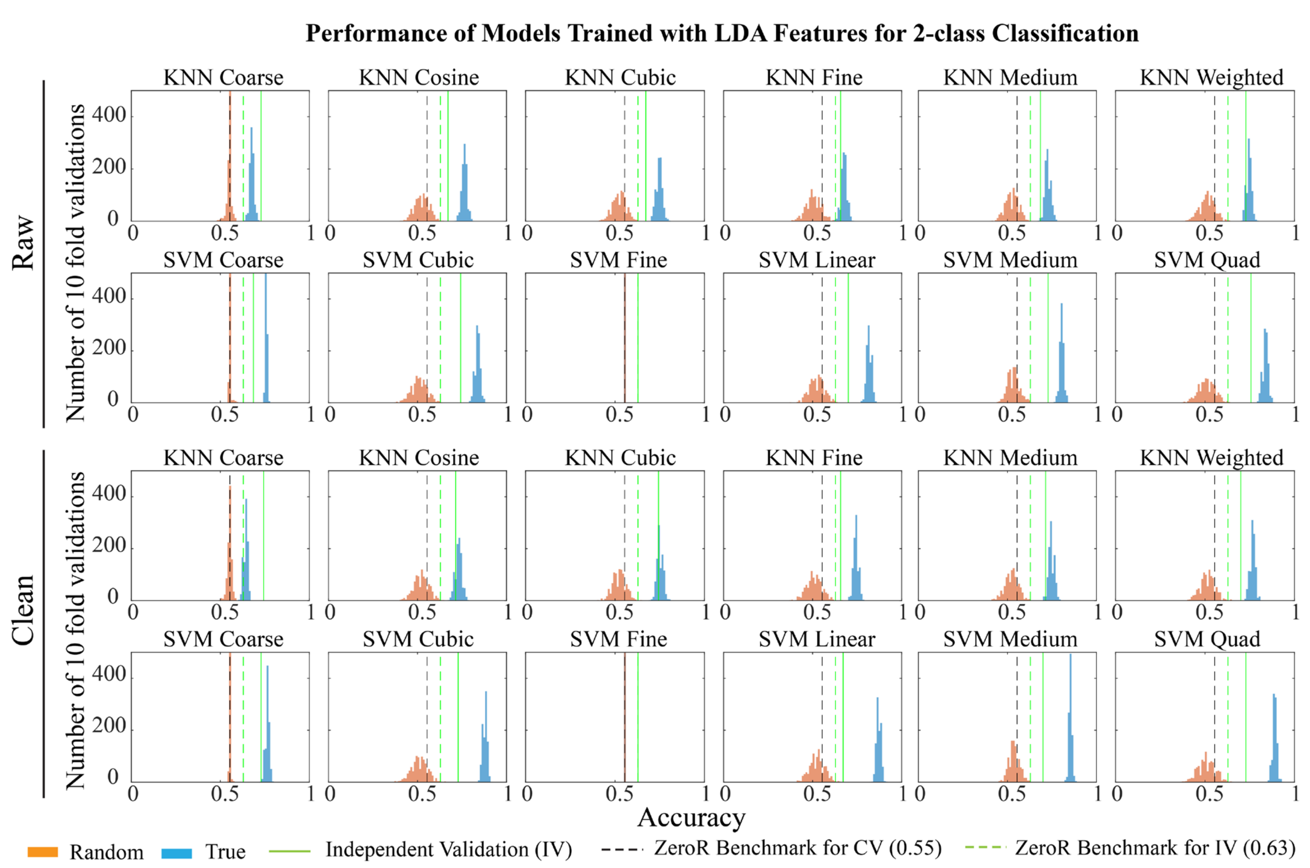


Figure 1. 1000 trials of CV accuracy vs randomly labeled data and IV accuracy for LDA.

Selection	Features (Raw/Clean)		Accuracy (CV/IV)	
Stats	98	82	0.68 ± 0.05	0.70 ± 0.05
LDA	208	224	0.70 ± 0.04	0.76 ± 0.09
FSFS	9	13	0.71 ± 0.06	0.67 ± 0.05
BSFS	181	218	0.75 ± 0.09	0.56 ± 0.06
PCA	132	132	0.58 ± 0.05	0.48 ± 0.10

Table 1. Numbers of features selected and average accuracies for 2-class classification.

Results and Discussion

Rank	Selection	Model	Setting	Artifact	Demo	Accuracy	IV Rank
1	BSFS	SVM	Quadratic	Clean	Yes	0.85	87
2	BSFS	SVM	Cubic	Clean	Yes	0.85	104
3	LDA	SVM	Quadratic	Clean	Yes	0.84	13
4	LDA	SVM	Cubic	Clean	Yes	0.84	25
5	BSFS	SVM	Quadratic	Clean	No	0.82	72
6	LDA	SVM	Quadratic	Clean	No	0.82	5
7	BSFS	SVM	Cubic	Clean	No	0.81	98
8	LDA	SVM	Cubic	Clean	No	0.81	5
9	BSFS	SVM	Linear	Clean	Yes	0.80	37

Table 2. Best performing 3-class models by CV accuracy

Rank	Selection	Model	Setting	Artifact	Demo	Accuracy	CV Rank
1	FSFS	SVM	Coarse Gaussian	Clean	Yes	0.71	54
1	LDA	SVM	Linear	Clean	No	0.71	10
1	LDA	SVM	Medium Gaussian	Clean	No	0.71	15
4	FSFS	SVM	Medium Gaussian	Clean	Yes	0.70	33
5	LDA	SVM	Linear	Clean	Yes	0.69	12
5	LDA	SVM	Quad	Clean	No	0.69	6
5	LDA	SVM	Cubic	Clean	No	0.69	8

Table 3. Best performing 3-class models by IV accuracy

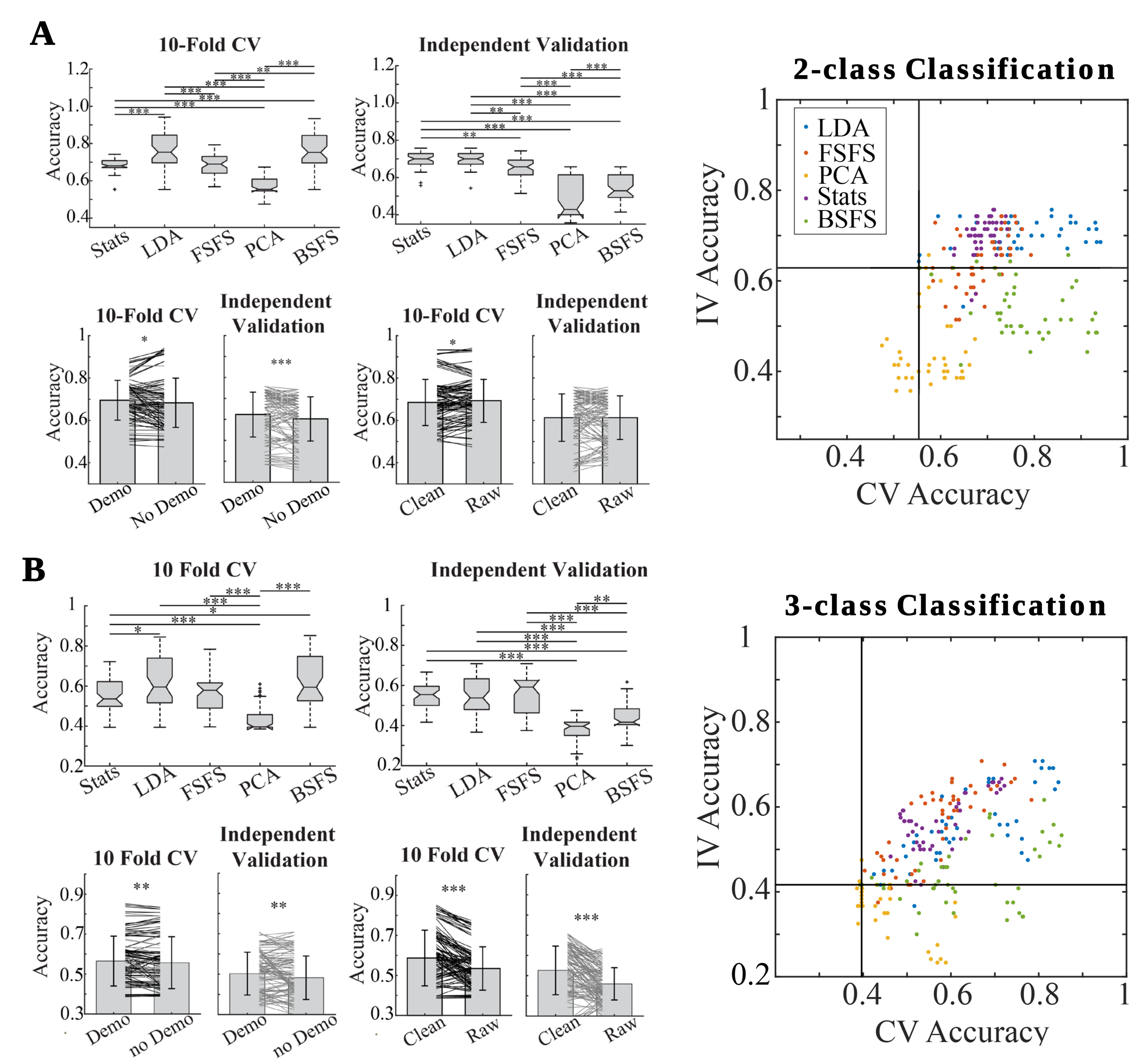


Figure 2. Summary of the performance of models. A. CV and IV 2-class accuracies for all selection methods. **B.** CV and IV 3-class accuracies for all selection methods. Dark lines indicate significant difference in two sample K-S test at 10^{-10} significance level. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, One way ANOVA and post-hoc/Signed-rank test.)

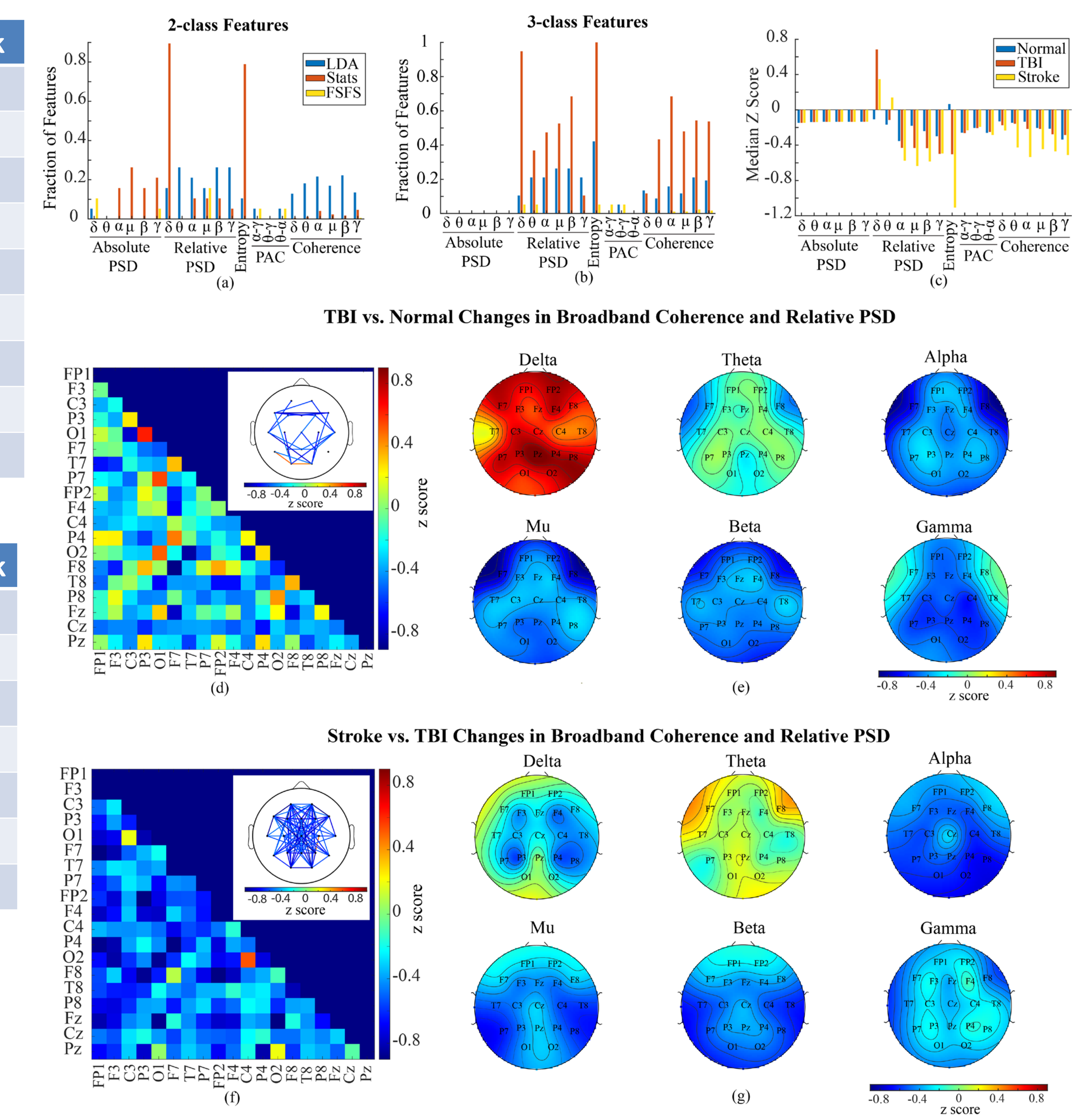


Figure 3. **qEEG feature changes in TBI subjects.** Features selected (a, b), feature changes from normal baseline (c), coherence changes between EEG contacts (d,f),and relative PSD changes (e,g).

Conclusion

This study demonstrates that ML models built upon qEEG features and demographic information extracted from existing public databases could distinguish between TBI and normal patients with up to 0.94 accuracy and 0.94 sensitivity in CV and 0.76 accuracy and 0.80 sensitivity in IV. With the addition of a cohort of stroke patients, these models were able to outperform a theoretical model that could only detect changes between normal and abnormal EEGs. Feature selection method appears to play the most important role in the performance of models. Our study shows LDA feature selection method outperformed all other methods, reflected by the observation that best performing models in CV and IV for both two- and three-class classification were predominantly based on features selected by LDA. In general, including demographic information in the input feature can significantly increase the performance of models, but to a limited degree. In line with prior qEEG study on TBI patients, coherence and relative spectral density were two major parameters changed from normal to TBI. Coherence change varied among channel pairs with reduction more predominant. Relative PSD demonstrated a global increase in low frequency delta power and decrease in higher frequency (alpha, mu, beta, and gamma) power. These results suggest EEG ML can potentially be used in the detection or monitoring of TBI in clinic.