medRxiv preprint doi: https://doi.org/10.1101/2022.10.22.22281402; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Post Traumatic Seizure Classification with Missing Data using Multimodal Machine Learning on dMRI, EEG, and fMRI

Md Navid Akbar, Sebastian F. Ruf, Ashutosh Singh, Razieh Faghihpirayesh, Rachael Garner, Alexis Bennett, Celina Alba, Marianna La Rocca, Tales Imbiriba, Deniz Erdoğmuş, and Dominique Duncan

***Abstract*— Late post-traumatic seizure (LPTS) is a complication of traumatic brain injury (TBI), which can lead to a potentially lifelong condition of post-traumatic epilepsy (PTE). Currently, the patho-mechanism that induces epileptogenesis in TBI subjects is unclear. As such, the epilepsy community strives to identify which TBI subjects will develop epilepsy and find potential biomarkers. To that end, this study collects longitudinal multimodal data from TBI subjects at multiple participating institutes. A supervised, binary classification task is formed with data from the LPTS versus no LPTS subjects. Missing modalities in certain subjects is handled in two ways. First, we extend a graphical model based Bayesian estimator to directly classify subjects with missing modality, and second, we investigate standard imputation techniques. The multimodal information is then combined, following several fusion and dimensionality reduction techniques found in literature, and eventually fitted to a kernel- or a tree-based classifier. For this fusion, we propose two new algorithms: recursive elimination of correlated components (RECC) which filters information based on correlation, and information decomposition and selective fusion (IDSF) which meaningfully recombines information from decomposed multimodal features. Based on the cross-validated area under the curve (AUC) score, we find the proposed IDSF algorithm provides the best performance. Finally, following statistical analyses of the frequently selected features, we recommend alterations in inferior temporal gyrus as a potential biomarker.**

***Index Terms*— seizure classification, multimodal fusion, machine learning, missing data, multivariate information, biomarker identification.**

I. Introduction

Post-traumatic epilepsy (PTE) is a form of acquired epilepsy that results from a traumatic brain injury (TBI) caused by an external force, for example a fall or a motor vehicle accident. Risk factors for PTE include occurrence of posttraumatic seizures (PTS) such as late PTS (LPTS, 1 week post-injury), because they carry a high risk of future seizures [1]. Effective pharmacological treatments in the prevention or treatment of symptomatic PTE seizures does not currently exist, therefore it is essential for the epilepsy community to find potential biomarkers, and techniques to identify which TBI subjects will likely develop LPTS [2]. TBI causes widespread damage to both functional and structural features and to properly understand the effect of TBI, we need to consider multiple modalities [3]. Currently, the Epilepsy Bioinformatics Study for Antiepileptogenic Therapy (EpiBioS4Rx) [4] is pioneering this work with the aim to design and perform preclinical trials of antiepileptogenic therapies, using multimodal information, followed by future planning of clinical trials.

For classification tasks utilizing multiple modality information, three major approaches are found in literature. They are, namely: late, intermediate, and early fusion. Late fusion involves combining predictions from multiple different base learners (Base), or heterogeneous ensemble classification, and it has been shown to be an efficient way of improving predictive accuracy in machine learning (ML) [5]. As a form of late fusion, stacking involves learning how to best combine the classifiers, in addition to combination using simple statistics, such as voting or averaging [6]. Džeroski et al. claimed that heterogeneous base learners, stacked with a final meta classifier, can perform comparably, if not better, to the best of the individual classifiers [6]. In the second major approach, intermediate fusion, features or information extracted from the individual modalities are aggregated, and subsequently fed to a single classifier. Since fusion in this method occurs in the feature space, it is often more interpretable, and also allows the application of multivariate information theoretic fusion [7]. The final approach, early fusion, involves representing the raw individual modality information collectively in a common space. From this aggregated information, an ML pipeline can then extract relevant features and perform the desired classification. Early fusion can be challenging to interpret, since meaningful aggregation of different types of modalities (e.g. spatial and temporal) is not always straightforward [8]. Past studies involving magnetic resonance imaging (MRI) and electroencephalography (EEG) have found intermediate and late fusion to be less affected by noise, misregistration, and to have lower complexity; compared to early fusion [8], [9].

While acquiring data from all modalities for each subject is ideally desirable, it is often not feasible. This creates a challenge of handling such missing modality information. Missing data has historically been dealt with imputation: using the existing data to make estimates for the missing values. Imputation techniques generally take either the form of: a univariate approach that considers each missing fea-

This project was funded by the National Institute of Neurological Disorders and Stroke (NINDS) of the National Institutes of Health (NIH) under award number R01NS111744.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.22.22281402; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

ture/ modality in isolation across all subjects, such as mean imputation; or a multivariate approach that considers the other available features/ modalities, which could be within subject or across all subjects, such as k-nearest neighbor (kNN) imputation [10]. Following imputation, standard fusion and classification pipelines can be implemented. Aside from conventional imputation, contemporary approaches include jointly performing data imputation and self-representation learning [11], or graphical model estimators that marginalize missing attributes [12].

In this work, we undertake a LPTS versus no LPTS binary classification task, in selected subjects from EpiBioS4Rx. In Section II, we begin by elaborating our data collection and preparation strategies. We considered three data modalities: functional MRI (fMRI), diffusion-weighted MRI (dMRI), and EEG. To address missing modality in certain subjects, first we extend a graphical model based Bayesian estimator, and second we investigate several univariate and multivariate imputation techniques. For combining the resultant multimodal information, aligned with the findings from relevant previous studies as discussed earlier, our work is based on intermediate and late fusion. For intermediate fusion, we propose two novel algorithms: one that refines multivariate information based on correlation, and the other selectively combines information following decomposition of the complete data. Finally, we search for potential biomarkers, following statistical comparisons of the most frequently chosen features between the two LPTS groups. The remainder of the paper is organized as follows: Section III presents our empirical findings, Section IV summarizes our discussion, and Section V concludes our work.

II. Methodology

A. Data Collection and Preparation

1) Data Acquisition: This work was approved by the UCLA Institutional Review Board (IRB$\#16-001576$) and the local review boards at each EpiBioS4Rx Study Group institution. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

According to the EpiBioS4Rx protocol [4], moderate-to-severe TBI subjects with frontal and/or temporal hemorrhagic contusion and Glasgow Coma Scale (GCS) score between $3-13$ were eligible for enrollment. A total of 48 subjects (36 male, 12 female; age = 42.1 ± 19.3 years; GCS = 7.8 ± 4.44) were chosen for this work. Of these subjects, $17(35\%)$ experienced at least one LPTS, whereas the remaining 31 did not experience any for the entire two year follow-up period, which is the minimum duration needed to identify about $80-90\%$ of the subjects who will eventually develop PTE [13]. Even though the modalities of interest for this work are dMRI, EEG, and fMRI, the acquired data from the subjects included other MRI sequences such as T1-weighted, T2-weighted, T2-weighted fluid attenuated inversion recovery (T2-FLAIR), etc. The following three sections will summarize the preprocesing steps involved in preparing our three data modalities of interest for the extraction of relevant features.

2) dMRI Preprocessing: dMRI scans corresponding to multiple diffusion gradient values and directions were collected. Acquired dMRI data was processed in the Oxford FMRIB Software Library (FSL) [14]), to estimate diffusion tensor imaging (DTI) parameters, such as fractional anisotropy (FA) maps, in subject-specific spaces. Since FA features have been reported in earlier work to be promising in characterizing PTS [15], [16], we focused on extracting and utilizing FA features in this work.

The collected individual subject FA images were then transformed to a common Montreal Neurological Institute (MNI) space by registering to a standard HCP 1065 DTI FA template, following Pipeline 1 in [16]. Finally, tract based spatial statistic [17] of each registered image was carried out using the mask and distance map obtained from the standard template, to extract mean FA values along 63 white matter (WM) tracts and bundles obtained from the JHU-DTI atlas [18]. These mean FA values are recorded as the dMRI features (x_d).

Exclusion Criteria: Of the 45 subjects analyzed, 41 subjects, meeting the two-year follow-up requirement, were chosen for this work. From the chosen, $14(34\%)$ experienced LPTS, whereas the remaining 27 did not.

3) EEG Preprocessing: Epileptiform abnormalities (EAs) such as seizures, periodic discharges (PDs), and abnormal rhythmic delta activity (ARDA), are potential biomarkers of epileptogenesis [19], [20]. Kim et al. [21] showed that the presence of EAs in the EEG signal during the acute period following TBI independently predicted PTE in the first year post injury. Clinicians annotated the obtained EEG to denote the presence of three such EAs (Seizure, PD, and ARDA), which formed our EEG features (x_e).

Exclusion Criteria: EEG recordings from 10 subjects were selected, following review by EEG experts. Of them, $7(70\%)$ developed LPTS, while the other $3(30\%)$ did not.

4) fMRI Preprocessing: To incorporate lesion information in the rs-fMRI preprocessing, we manually segmented damaged brain tissue into parenchymal contusions and brain edema with ITK-SNAP [22] from the acquired 3D T2-FLAIR scans. Brain contusion was defined as a lesion with abnormal signal intensity and hemorrhagic volume > 1 ml, and brain edema was defined as a region surrounding or in the proximity of the contusion with hyperintense signal compared with the WM signal on T2-FLAIR images [23]. Each segmentation was carried out by a student research assistant and reviewed for clinical accuracy by one of two medical physicians with research experience in neuroradiology. For each TBI subject, we obtained a 3D lesion mask including contusion and edema. For the purposes of this work, the lesion incorporation served to obtain a more accurate preprocessing pipeline, in which brain alterations related to TBI were not mislabeled as brain tissues such as WM, cerebrospinal fluid (CSF) and grey matter (GM). Therefore, we did not consider the edema and contusion separately within each 3D mask. Then, to use the lesion masks in the rs-fMRI preprocessing pipelines, we performed affine registration on each mask using the MNI 152 template with the Linear Registration Tool (FLIRT) of FSL [14]. Affine transformations were used because they allowed us to better maintain the morphological characteristics of the lesions while at the same time obtaining robust registrations.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.22.22281402; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The preprocessing used a modification of the unified segmentation normalization of SPM12 [24] as accessed through the CONN toolbox [25]. The literature on normalization suggests that SPM12's unified segmentation normalization outperforms other pipelines when applied to simulated lesions [26]. However further work [27] showed that to achieve the best performance, information about the lesion should be included, motivating the inclusion of the lesion mask in preprocessing.

The fMRI preprocessing pipeline incorporates information about the lesion by performing SPM12's unified segmentation normalization using a lesion modified Tissue Probability Map (TPM). As discussed in [27], modifying the TPM performs implicit cost function normalization because it ultimately causes SPM to ignore the lesion areas during normalization. The TPM normally provides a prior probability of a given voxel belonging to one of 6 tissue classes; GM, WM, CSF, skull, soft tissue, and other [28]. SPM12's default TPM was augmented for each subject with a 7th tissue class which corresponds to the MNI lesion mask for that subject, thereby setting the prior probability of the existence of GM, WM, or CSF in the lesion areas to 0. A new function in the CONN toolbox, `conn_createtpm`, was used to create the modified TPM. Finally, using the AAL atlas, the strengths of the positive (x_{f-p}), negative (x_{f-n}), and overall (x_{f-o}) connections were recorded as the fMRI features [29].

Exclusion Criteria: Of the 44 subjects analyzed, 8 were excluded because the preprocessing failed or the subject's imaging exhibited significant motion (less than $\frac{1}{3}$ of the scan was valid). From the remaining, scans corresponding to 32 subjects, meeting the two-year follow-up requirement, were chosen. Out of these, 12 developed LPTS (38%), whereas the remaining 20 (62%) did not.

B. Data Distribution and Preparation

1) Data Distribution by Modality and Train-Test Partitions: While the goal of the study was to acquire all three modalities of interest (x_d, x_e, x_f) for each subject, all subjects did not have all modalities. Table I shows the number of subjects for each combination of modalities, as well as the relative number of subjects within each modality configuration. Out of the included total $N = 48$ subjects, only seven had information available from all three modalities. Of the other 41 subjects, 21 had a single missing modality, while the remaining 20 had two missing modalities.

From the total N subjects, we then created train and test partitions, following a five-fold cross-validation (CV). Any missing modalities were handled using methods described in Section II-B.2. We concatenated all D features (166×3 for the three fMRI strengths, 63 for dMRI, 3 for EEG) along the columns of $\tilde{X}_i \in \mathbb{R}^{1\times D}$, for $i \in n$ samples in each CV train fold. Each $\tilde{\mathbf{X}}_i$, representing an individual subject, is then stacked vertically as rows to give the training partition ($\mathbf{X}_{\text{train}}$) for each CV round

$$\mathbf{X}_{\text{train}} = \left[\tilde{X}_1,\ \tilde{X}_2,\ \ldots\ ,\tilde{X}_n\right]^\top.$$

Each test partition ($\mathbf{X}_{\text{test}}$) is constructed in a similar manner with $m = N - n$ samples. The complete feature set, including the train and the test sets, is denoted by $\mathbf{X} = \left[\mathbf{X}_{\text{train}}, \mathbf{X}_{\text{test}}\right]^\top$,

TABLE I
DISTRIBUTION OF SUBJECTS BY THE DIFFERENT AVAILABLE MODALITIES AND LPTS LABELS.

	Subjects	fMRI	dMRI	EEG	LPTS	No LPTS
	5	✓			1	4
	13		✓		2	11
	2			✓	2	0
	20	✓	✓		7	13
	1		✓	✓	1	0
	7	✓	✓	✓	4	3
Total:	48	32	41	10	17	31

where $\mathbf{X} \in \mathbb{R}^{N\times D}$. Similarly, the train-test partitions for the complete set of LPTS labels $\mathbf{y} \in \mathbb{R}^{N\times 1}$, corresponding to any CV round, could be expressed as $\mathbf{y} = \left[\mathbf{y}_{\text{train}}, \mathbf{y}_{\text{test}}\right]^\top$.

2) Missing Data Imputation Technique: Most classification techniques in literature cannot deal with missing data. In order to establish an individual modality baselines and utilize several relevant classification pipelines, we investigated imputing the missing modality features following four standard techniques, namely: mean, median, kNN, and iterative. The mean and median imputation strategies replace each missing feature in $\tilde{\mathbf{X}}_i$ with the mean and median, respectively, of the non-missing values in the corresponding feature column of $\mathbf{X}_{\text{train}}$. For the kNN imputation, each missing feature in $\tilde{\mathbf{X}}_i$ is imputed using the mean value of the feature present in a chosen number of nearest neighbors s_{ne} (in terms of Euclidean distance among the non-missing features) of $\tilde{\mathbf{X}}_i$ within $\mathbf{X}_{\text{train}}$ [30]. For the iterative imputation, we imputed the missing values in $\tilde{\mathbf{X}}_i$ by modeling d features with missing values as a function of the other $D - d$ features in a round-robin fashion, for a given number of iterations. For all imputation techniques, we used information from $\mathbf{X}_{\text{train}}$ only (in an unsupervised learning setting), and not from $\mathbf{y}_{\text{train}}$.

C. Model Pipeline and Evaluation

1) Dimensionality Reduction (DR) and Classifiers: Considering $\mathbf{X}_{\text{train}}$ as input, the feature dimension $D \gg n$. Aside, by inclusion of all features, we risk hurting the final classification performance by adding noise, and over-fitting by increasing complexity [31]. It is thus recommended to reduce the dimensionality of such high dimensional input. Thus, a standard DR technique is cascaded at the start of each classifier for all fusion approaches, aside the intermediate fusion techniques (Other) outlined in Sections II-G.2-II-G.7, which themselves serve to reduce dimensions. The three standard reduction procedures adopted in this work are: principal component analysis (PCA) [32], and K-best features: chosen by maximizing either the χ^2-test score (χ^2_{score}) or the Fischer score (f_{score}), between $\mathbf{X}_{\text{train}}$ and $\mathbf{y}_{\text{train}}$ [33].

In all classification tasks, whether as an interim individual learner or as the final estimator, both SVM and AdaBoost classifiers were investigated. The choice of these two classifiers was aligned with previous studies, which have found both SVM [34], [35] and AdaBoost [16] to be suitable for classification tasks involving neuroimaging features. For SVM, both linear and radial basis function (RBF) kernel variants were explored.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.22.22281402; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

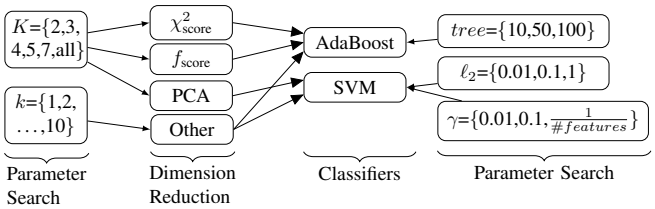

Fig. 1. Model pipeline showing the different dimension reduction techniques and the classifiers, along with their parameter search spaces.

The top K or k features/components to be chosen, for either one of the standard techniques or the intermediate fusion techniques, were varied in the search space as shown in Fig. 1. The only difference in implementation was: different values for K were chosen per CV fold for the standard reduction techniques, following a nested 5-fold CV within the train set, whereas the same value for k was utilized (for computational considerations) for all outer CV folds involving the intermediate reduction techniques. For the classifiers, the search space for SVM RBF's gamma (γ), L-2 regularization (ℓ_2), and number of trees ($tree$) in AdaBoost can also be seen in Fig. 1. All models were trained and tested in Python 3.7.4, with support from scikit-learn [33] version 1.0.2, CCA-Zoo [36], and MMIDimReduction [37].[1]

2) Scoring Metrics: For evaluating models on the test partitions in each CV fold, we use mean area under the receiver operating characteristic (ROC) curve (AUC) as the primary metric. Hanley et. al. demonstrated AUC to be an effective measure of accuracy, with a meaningful interpretation in medical diagnosis [38]. We thus used AUC to rank the models. Aside, AUC helps choose a desired operating point (sensitivity, specificity) in the ROC curve, where

$$\text{sensitivity} = \frac{TP}{TP+FN}, \text{ and specificity} = \frac{TN}{FP+TN},$$

where TP, FN, TN, and FP are the true positive, false negative, true negative, and false positive, respectively. Two techniques have been suggested in literature [39] to choose the best operating point (for equal weights to sensitivity and specificity), including the highest

$$\text{Youden index} = \text{sensitivity} + \text{specificity} - 1, \tag{1}$$

and the lowest Euclidean distance from the upper left corner (0,1) of the ROC curve, as

$$d_{\text{UL}} = \sqrt{(1-\text{sensitivity})^2 + (1-\text{specificity})^2}. \tag{2}$$

As a secondary metric to assess model performance on this imbalanced dataset, but not to rank them, we used weighted F1-score as given by

$$F1 = \sum_{i=1}^{2} \frac{m_i}{m} \frac{2*TP}{2*TP+FP+FN}, \tag{3}$$

where m_i is the number of samples belonging to the i-th class, in the current test fold.

D. No Imputation: Naive Bayes (NB) Late Fusion

In this section, we will extend the standard NB classifier [12] to our graphical model to handle missing modalities in the original data. As seen in Fig. 2, given the LPTS label (y), we hypothesized that the resulting physical state of the subject

[1]The code used to generate the results will soon be made publicly available.

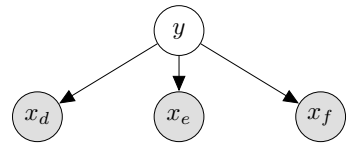

Fig. 2. Graphical model depicting the hypothesized relation of the observed modalities (shaded circles), with LPTS label as the latent state (clear circle).

was captured by the different observed modalities (x_d, x_e, x_f). The purpose of multimodal fusion with the different modalities was to maximize the probability of detecting the true LPTS label, given the collected evidence. In the NB approach, features from each modality were assumed to be conditionally independent of each other, such that

$$p(x_d, x_e, x_f|y) = p(x_d|y)p(x_f|y)p(x_e|y). \tag{4}$$

Under the conditional independence assumption, it is possible to integrate out the contribution of any missing modalities, i.e., by marginalizing over all possible values of x_j given y, where $j \in M$ and $M = \{d, e, f\}$. This formulation can be directly extended to missing modality scenario. For instance, suppose x_e is the missing modality, Eq. (4) becomes $p(x_d, x_f|y)$ which is equivalent of marginalizing over all possible values of x_e given y, leading to:

$$p(x_d, x_f|y) = p(x_d|y)\left(\int p(x_e|y)dx_e\right)p(x_f|y) \tag{5}$$

where $\int p(x_e|y)dx_e = 1$.

The maximum likelihood estimation of choosing the predicted label y of a given subject is then taken to follow

$$y = \arg\max_{l\in\{0,1\}} \frac{p(x_d, x_e, x_f|y=l)p(y=l)}{p(x_d, x_e, x_f)}, \tag{6}$$

where $l = 0$ (no LPTS), $l = 1$ (LPTS), and

$$p(x_j|y) = \frac{p(y|x_j)p(x_j)}{p(y)} \quad \forall j \in M. \tag{7}$$

Since $p(x_j)$ and $p(x_d, x_e, x_f)$ are not functions of y, the optimization objective in (6) is equivalent to

$$y = \arg\max_{l\in\{0,1\}} p(y=l) \prod_{j\in M} \frac{p(y=l|x_j)}{p(y=l)}. \tag{8}$$

In this work, output from each individual modality base learner (elaborated in Section II-E) was used to approximate $p(y|x_i)$, and $p(y=l)$ is estimated from the train set.

E. With Imputation: No Fusion

Here, we will summarize how we developed individual modality base learners, using the data with imputation and without any explicit fusion (disregarding the cross-modality information lookup during imputation). The imputed features from each j-th ($j \in M$) modality were passed through a standard reduction technique and then subsequently fitted to a classifier, as outlined earlier in Fig. 1.

F. With Imputation: Late Fusion

In this section, we will discuss the implementation of the late fusion ensemble classification on the imputed data. We begin by collecting the probabilities $p(y_j)$ corresponding to the labels

$$y_j = l \;\; \forall\, l \in \{0, 1\}$$

medRxiv preprint doi: https://doi.org/10.1101/2022.10.22.22281402; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

from each j-th modality classifier (base learner), as introduced in Section II-E. We then summed these probabilities for $l = 0$ and $l = 1$, and utilized maximum likelihood to estimate

$$y = \arg\max_{\ell \in \{0,1\}} \sum_{j \in M} p(y_j = l). \tag{9}$$

This manner of probabilistic fusion (soft) takes into account the confidence in the predictions of the individual base learners. This is in contrast to another comparable implementation, where each learner directly estimates the output label, and the fusion ultimately involves only majority voting (hard) among the estimated labels [5].

G. With Imputation: Intermediate Fusion

In this section, we will discuss the implementations of several existing intermediate fusion techniques, as well as propose two novel techniques at the end.

1) Feature Union: In this approach, all D features from the three modalities were concatenated to give $\mathbf{X}_{\text{train}} \in \mathbb{R}^{n \times D}$. Subsequently, $\mathbf{X}_{\text{train}}$ was passed through a standard reduction technique to extract K features, and then fitted to a classifier.

2) Canonical Correlation Analysis (CCA): We hypothesized that by looking in the directions of the maximal correlations of the input modalities, we would be able to extract information about their shared space, which might help the classification task, since the y was assumed to be the underlying common latent state in Fig. 2. In CCA, two sets of variables are linearly combined to obtain canonical variates such that the correlation between the linear combinations is maximized [32]. This technique was thus utilized to combine our input modalities in a three-fold combination: (x_d, x_e); (x_d, x_f); and (x_e, x_f).

For mathematical formulation, let us assume the combination (x_d, x_f). If we can express two linear functions $u = a^\top x_f$ and $v = b^\top x_d$, and if the covariance [32]

$$\text{Cov}(x_d, x_f) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \tag{10}$$

it can be shown that their correlation is maximized when

$$\begin{bmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = 0, \tag{11}$$

where the eigenvalues $\lambda \in \{\rho_1 \geq \ldots \geq \rho_{s+t}\}$, in descending order. The maximum correlation was obtained with $\lambda = \rho_1$, and it provided our first pair of variates $u_1 = a_1^\top x_f$ and $v_1 = b_1^\top x_d$. From the obtained D such pair of variates, we then selected k_{CCA} pairs to form $\mathbf{Z}_{\text{CCA}} \in \mathbb{R}^{n \times k_{\text{CCA}}}$ ($k_{\text{CCA}} \leq D$), and fitted $\mathbf{Z}_{\text{CCA}}$ to a final classifier.

3) Generalized Canonical Correlation Analysis (GCCA): CCA can only linearly combine two sets of variables. GCCA, however, can combine more than two variables [40]. In our case, GCCA combined (x_d, x_e, x_f) directly. GCCA attempts to solve the following optimization problem

$$\begin{aligned} \max_{\alpha_1, \ldots, \alpha_J} \quad & \sum_{j,q=1, j \neq k}^{J} c_{jq} g\big(\text{corr}(\alpha_j^\top \mathbf{X}_j, \alpha_q^\top \mathbf{X}_q)\big) \\ \text{s.t.} \quad & \text{Var}(\alpha_j^\top \mathbf{X}_j) = 1, \quad j = 1, \ldots, J, \end{aligned} \tag{12}$$

where $\text{corr}(\cdot)$ is the Pearson correlation function, $g(\cdot)$ is the identity function, $c_{jq} = 1$ represents the connections between modality feature matrices $\{\mathbf{X}_d, \mathbf{X}_e, \mathbf{X}_f\}$, and J=3.

Similar to CCA, we then selected k_{GCCA} pairs of variates from the total D pairs, to give $\mathbf{Z}_{\text{GCCA}} \in \mathbb{R}^{n \times k_{\text{GCCA}}}$ ($k_{\text{GCCA}} \leq D$), and fitted $\mathbf{Z}_{\text{GCCA}}$ to a final classifier.

4) Sequential Feature Selection (SFS): For selection of features from such high D dimensional data, standard univariate feature selection techniques ignore the mutual information among features. Multivariate feature selection techniques, however, consider a group of features in its entirety. Unfortunately, searching for the globally optimal subset with exhaustive search is $O(2^n)$, and can be computationally intractable. As a result, it is common practice to resort to algorithms striving to obtain a locally optimal, but perhaps globally sub-optimal, feature set with a lower complexity [34]. Wrapper method, a multivariate feature selection technique, attempts to find a set of highly important features by fitting a particular classifier on all features in a nested CV within the train set [31]. Wrapper methods have been demonstrated to provide superior performance (compared to filter methods in feature selection) in classification tasks involving high-dimensional neuroimaging data with limited samples [34].

In SFS, a multivariate wrapper technique, we greedily choose features for classification, either by forward selection or by backward elimination. For the first round in forward selection, based on the score of the wrapper classifier, the most informative feature is added. In the case of backward selection, the least informative feature is removed. Forward selection thus involves a bottom-up search strategy, which begins with an empty set, and during each iteration, a new feature is added to the current set so the loss function is reduced [34]. On the contrary, backward elimination, follows a top-down approach, starting with the complete set D, features are removed one at a time such that the reduction in performance is kept at a minimum [34]. Choosing either forward or backward selection, we retain k_{SFS} features. Selected features are collected as $\mathbf{Z}_{\text{SFS}} \in \mathbb{R}^{n \times k_{\text{SFS}}}$ ($k_{\text{SFS}} \leq D$), and subsequently fitted to a final classifier. SFS has an average complexity of O(n^2).

5) Stochastic Mutual Information Gradient (SMIG): In this method, we aim to learn a feature transformation network $\phi^\star : \mathbb{R}^{n \times D} \mapsto \mathbb{R}^{n \times k_{\text{SMIG}}}$ such that the high $n \times D$-dimensional input feature space is mapped to a lower $n \times k_{\text{SMIG}}$-dimensional transformed feature space [37]. This mapping is done while maximizing the mutual information between the transformed set and the train labels $\mathbf{y}_{\text{train}}$ as

$$\phi^\star = \arg\max_{\phi \in \Omega} \{I(\mathbf{Z}_{\text{SMIG}}, \mathbf{y}_{\text{train}})\}, \tag{13}$$

where $I(\cdot)$ is the mutual information function, $\mathbf{Z}_{\text{SMIG}}$ contains the transformed training samples, and Ω denotes the function space for possible feature mappings ϕ [37]. The parameters of this $\phi^\star(\cdot)$ network is updated iteratively by a technique known as SMIG [37].

6) Recursive Elimination of Correlated Components (RECC): In CCA, each of the k_{CCA} pairs of canonical variates chosen is uncorrelated to each other, since the pairs maintain orthogonality among themselves [32]. However, there is no such imposed constraint on some of the other DR techniques like SFS or SMIG. As such, there exists a possibility that the chosen set of k features/projections (components) will likely contain a set of components, which are highly correlated with

medRxiv preprint doi: https://doi.org/10.1101/2022.10.22.22281402; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Algorithm 1: RECC for a given selection algorithm

Input : Pearson correlation coefficient threshold ρ_{thresh}, column components $\mathbf{Z}_{\text{sorted}} \in \mathbb{R}^{n \times k_1}$ (sorted by decreasing order of importance by the selection algorithm, $k_1 \leq D$)
Output: column components $\mathbf{Z}_{\text{RECC}} \in \mathbb{R}^{n \times k_2}$ ($k_2 \leq D$)

```
Initialize j = k_1, Z_RECC = Z_sorted
while j > 1 do
    for i = 1 ... (j − 1) do
        ρ ← corr(Z_sorted[:, i], Z_sorted[:, j])
        if ρ ≥ ρ_thresh then
            Z_RECC ← Z_RECC - Z_RECC[:, j]
        end
    end
    Decrement j
end
return Z_RECC
```

Algorithm 2: IDSF for one decomposition algorithm and one selection algorithm prepared by RECC

Input : Column components $\mathbf{Z}_{\text{shared}} \in \mathbb{R}^{n \times k_1}$ ($k_1 \leq D$, decomposed information shared by modalities obtained using algorithm-1), column components $\mathbf{Z}_{\text{unique}} \in \mathbb{R}^{n \times k_2}$ ($k_2 \leq D$, decomposed information present uniquely within modalities estimated by applying RECC on algorithm-2)
Output: fused components $\mathbf{Z}_{\text{IDSF}} \in \mathbb{R}^{n \times k_3}$, where $k_3 \leq D$

```
Initialize Z_IDSF = Z_shared
for j = 1 ... k_2 do
    Z_IDSF ← Z_IDSF + Z_unique[:, j]
end
return Z_IDSF
```

each other. Inclusion of such components might negatively impact the classification performance [31].

To address this, we sought to remove the lesser important set of highly correlated components using the proposed technique, RECC, as outlined in Algorithm 1. We start with $\mathbf{Z}_{\text{sorted}}$ components sorted in decreasing order of importance, according to given selection algorithm. Starting with the least important component, if the Pearson correlation coefficient (ρ), of the latest component in comparison with any of the more important components, is $\geq \rho_{\text{thresh}}$, we drop that latest component from $\mathbf{Z}_{\text{RECC}}$. We continue this comparison, and elimination if applicable, until we have compared all the components in $\mathbf{Z}_{\text{RECC}}$, which is then fitted to the final classifier.

7) Information Decomposition and Selective Fusion (IDSF): Our earlier hypothesis was that by utilizing information from the shared space among the modalities, we would hope to observe a better classification performance, since the LPTS label y was also assumed to be a shared latent state. However, there exists the possibility that a fraction of the information, useful for classification, is present uniquely in a space not shared by the individual modalities [41]. This indicates that a classifier could potentially benefit from a combination of the shared information, as well as modality-specific unique information, obtained from the multiple modalities. This new hypothesis formed the basis of our proposed IDSF algorithm, where we fused the shared information (e.g. canonical variates from GCCA) with the estimated unique information (e.g. components selected by RECC, using a given selection algorithm like SFS). IDSF is elaborated in Algorithm 2.

III. Results

Based on our preliminary classification results with the four different imputation techniques, kNN imputation outperformed the other three. For this reason, all subsequent results in this section, wherever data was imputed, was done using the kNN imputer, with $s_{\text{ne}} = 1$.

The comparison of the mean AUC and the mean weighted F1 performances, on the test sets from all five CV folds, for the different fusion techniques (except IDSF) are summarized in Table II. The first row shows the classification performance of our extended NB estimator with late fusion, where the original input features (no imputation) were used to fit individual modality base learners. NB (AUC=0.710) outperformed the individual modality base learners (AUC=$\{0.664, 0.670, 0.619\}$), using imputed features without any fusion, and which form the next three rows of Table II.

Of the techniques using imputed features and fusion, we first list the performance of the soft late fusion estimator. This probabilistic estimator also performed better (AUC=0.756) than the individual modality base learners. It even performed better than both the late fusion NB estimator (no imputation), and the intermediate fusion (union of all modality features) classifier (AUC=0.736), in the subsequent row of Table II.

The last six rows of Table II correspond to the DR techniques involving intermediate fusion, as discussed earlier in Sections II-G.2-II-G.6. Among them, the first four rows presented existing techniques in literature, and the best performance (AUC=0.784) was achieved with CCA. Even though GCCA took into account all three modalities (x_d, x_e, x_{f-n}), compared to CCA's two (x_d, x_{f-p}), we noticed it lagged in performance (AUC=0.654) substantially compared to that of CCA. Of the two SFS implementations discussed (forward and backward), forward selection performed better (AUC=0.676), and was thus recorded. Similarly, between the two implementations of SMIG (linear and non-linear), linear performed better (AUC=0.699) and was recorded. Finally, the last two rows in Table II demonstrated the improvement in performances with our proposed RECC (AUC=$\{0.710, 0.753\}$), when applied on the original SFS and SMIG implementations, respectively. Aside, the best F1 score (0.753) was obtained with RECC-SMIG, indicating a potential choice for the best model, if F1 score was more important in the model selection.

To obtain a better understanding of how the choice of ρ_{thresh} affected the AUC performances of the RECC combinations corresponding to the last two rows in Table II, we plotted AUC vs. ρ_{thresh} in Fig. 3. The standard implementations of SFS and SMIG corresponded to $\rho_{\text{thresh}} = 1$. As we lowered ρ_{thresh}, we noticed the AUC performances to go up, until it peaked at $\rho_{\text{thresh}} = 0.5$ for both RECC implementations, and then started to go down. At $\rho_{\text{thresh}} = 0.15$, the performance of RECC-SMIG almost returned to the its SMIG baseline, whereas that of RECC-SFS remained higher than its SFS baseline.

In Table III, we recorded the best AUC performances for each combination of the two decomposition algorithms (CCA and GCCA) with the two selection algorithms (SFS and SMIG) prepared by RECC, using intermediate fusion with IDSF. We found the best IDSF performance (AUC=0.792), also the best among all tested models, was obtained with CCA and RECC-SFS. Just as seen earlier with CCA and GCCA, IDSF with CCA was superior to IDSF with GCCA. Interestingly, while individually both SMIG and RECC-SMIG

medRxiv preprint doi: https://doi.org/10.1101/2022.10.22.22281402; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

TABLE II

PERFORMANCE OF THE DIFFERENT DIMENSIONALITY REDUCTION TECHNIQUES, FOR DIFFERENT STAGES OF FUSION, WITH BOTH ORIGINAL (NO IMPUTATION) AND IMPUTED FEATURES.

Imputation	Fusion	Modalities	Method	Features/Components	Classifier	AUC	F1
No	Late	x_d, x_e, x_{f-o}	NB (f_{score})	Varied	Base	0.710	0.645
Yes	None	x_d	Base (χ^2_{score})	Varied	AdB	0.664	0.611
		x_e	Base (PCA)	Varied	SVM	0.670	0.639
		x_{f-o}	Base (PCA)	Varied	SVM	0.619	0.509
	Late	x_d, x_e, x_{f-o}	Soft Ensemble	Varied	Base	0.756	0.616
	Intermediate	x_d, x_e, x_{f-p}	Feature Union	All	AdB	0.736	0.651
		x_d, x_{f-p}	CCA	7	SVM	**0.784**	0.584
		x_d, x_e, x_{f-n}	GCCA	8	SVM	0.654	0.514
		x_d, x_e, x_{f-o}	SFS (fwd)	5	SVM	0.676	0.581
		x_d, x_e, x_{f-o}	SMIG (lin)	4	SVM	0.699	0.629
		x_d, x_e, x_{f-o}	RECC-SFS (fwd)	5 (Varied,ρ=0.5)	SVM	0.710	0.603
		x_d, x_e, x_{f-o}	RECC-SMIG (lin)	3 (Varied,ρ=0.5)	SVM	0.753	**0.753**

* lin=linear, fwd=forward, Varied=different number of features/components selected by grid search/ ρ_{thresh} in each CV fold, AdB=AdaBoost.

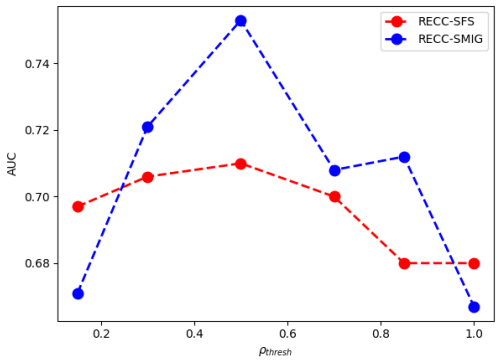

Fig. 3. AUC performance of RECC-SFS and RECC-SMIG combinations from the last two rows in Table II, as a variation of the Pearson coefficient threshold ρ_{thresh}, with SVM classifier and modalities $\boldsymbol{x_d, x_e, x_{f-o}}$.

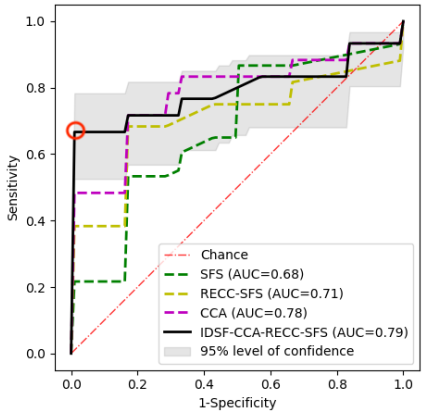

Fig. 4. ROC curve for the best performing model (solid black line) using IDSF-CCA-RECC-SFS. The ROC plots for the classification performance with SFS, RECC-SFS, and CCA are also shown: as purple, yellow, and green dotted lines, respectively. The best operating point (as per Youden index) is shown in a red circle. The classifier is SVM and information from all three modalities are used.

performed considerably better than SFS and RECC-SFS earlier, their IDSF performances were closely matched. Aside, while IDSF improved on the performances of the decomposition algorithms on all occasions, it did not necessarily do so over that of the selection algorithms (e.g. IDSF with GCCA and RECC-SMIG). Finally, the best F1 score (0.677) for IDSF was obtained with GCCA and RECC-SFS, indicating a potential runner-up for the best model, should F1 score be more important in the model selection.

To examine the effect of the proposed algorithms RECC and IDSF in terms of operating points, mean ROC curves (from the 5-fold CV) are plotted in Fig. 4 for the best model from Table III, as well as for the individual techniques that went into the best model from Table II. The variation of the best ROC curve (IDSF) is estimated by a binomial distribution approximation, for a 95% level of confidence [42]. We noticed that just like RECC-SFS improved on standard SFS, IDSF (AUC=0.79) improved on the performance over its two input techniques: CCA (AUC=0.78) and RECC-SFS (AUC=0.71). It also obtained the best operating point $(0.01, 0.67$: shown in a red circle), according to the highest Youden index (0.68) and the lowest d_{UL} (0.23).

In order to understand whether the features that resulted in the best classification performances could also be treated as potential biomarkers, we compared the distribution of the most frequently selected feature(s) between the two LPTS groups. These features were selected by the best algorithm on each data set (original and imputed), from the aggregate of all the five CV folds. Table IV presents this comparison, using a two-tailed non-parametric Mann Whitney U test. The features with unadjusted p-values less than 0.05 are reported in Table IV, along with their Bonferroni corrected p-values [43]. Table IV shows that for NB, using original features, inferior temporal gyrus (region-90) in x_{f-o} is significantly different between the two LPTS groups.

IV. DISCUSSION

We observed that in Fig. 4, the best AUC performance was observed with our proposed IDSF algorithm. While the improvement of IDSF over CCA was marginal (AUC=0.79 over AUC=0.78), the operating points of the two curves are quite distinct, as seen in Fig. 4. The choice of which point to operate in is often subject to clinical considerations, like if it is more important to correctly diagnose LPTS (sensitivity weighted higher), or no LPTS (specificity weighted higher), or the financial costs for correct and false diagnosis is the

medRxiv preprint doi: https://doi.org/10.1101/2022.10.22.22281402; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

TABLE III

PERFORMANCE OF IDSF (WITH ITS INPUTS) IN INTERMEDIATE FUSION WITH IMPUTED FEATURES.

Modalities	Shared	Comp.	Modalities	Unique	Components	Classifier	AUC	F1
x_d, x_{f-p}	CCA	7	x_d, x_e, x_{f-o}	RECC-SFS (fwd)	5 (Varied,ρ=0.5)	SVM	**0.792**	0.619
x_d, x_{f-p}	CCA	6	x_d, x_e, x_{f-o}	RECC-SMIG (lin)	1 (Varied,ρ=0.5)	SVM	0.783	0.584
x_d, x_{f-p}	GCCA	1	x_d, x_e, x_{f-o}	RECC-SFS (fwd)	5 (Varied,ρ=0.5)	AdB	0.663	**0.677**
x_d, x_{f-n}	GCCA	8	x_d, x_e, x_{f-o}	RECC-SMIG (lin)	4 (Varied,ρ=0.5)	AdB	0.670	0.642

* lin=linear, fwd=forward, Varied=different number of components were selected by ρ_{thresh} in each CV fold, AdB=AdaBoost.

TABLE IV

COMPARISON OF SELECTED FEATURES BETWEEN THE TWO LPTS GROUPS USING A TWO-TAILED MANN WHITNEY U TEST.

Imputation	Subjects	Model	Modality	Potential Biomarker	Unadjusted p-value	Comparisons	Corrected p-value
No	41	NB	x_d	EC-R	0.049	2	0.098
	32	NB	x_{f-o}	Region-90	0.025	2	**0.050**
Yes	48	IDSF-CCA-RECC-SFS	x_d	FX/ST-R	0.050	5	0.250
	48	IDSF-CCA-RECC-SFS	x_{f-o}	Region-48	0.049	5	0.245

* The p-values were first tested at a 5% level of significance ($p \leq 0.05$) and then corrected using Bonferroni correction, EC-R=external capsule right, region-90=inferior temporal gyrus, FX/ST-R=fornix/stria terminalis right, region-48=lingual gyrus right.

same (sensitivity=specificity) [39]. If either of the latter two aforementioned cases is desired, IDSF would perhaps emerge as the best model, for its operating point shown in red circle in Fig. 4, indicating the highest Youden index and the lowest d_{UL}. However, should the cost of a false negative diagnosis be higher, then the answer might not be as straightforward.

Aligned with our finding in Table IV, inferior temporal gyrus has been previously reported as a brain biomarker for early PTS (EPTS) and LPTS, following shape analysis [44]. It is also worth noting that alterations in the other regions, which we found to be statistically significant before the Bonferroni correction, have also been shown to be associated with PTS-related abnormalities: fornix/stria terminalis (FX/ST) [16], external capsule (EC) [45], and lingual gyri (regions-47,48) [46]. There is also the possibility that other potential biomarkers did not demonstrate a significant group level difference, owing to their variations within the individual level [47].

There have been several past studies involving multimodal ML classification of seizures, including classification from EEG and structural MRI [48], and from EEG and ECG [49]. However, to the best of our knowledge, ours is the first work that investigates multimodal ML classification of LPTS from dMRI, EEG, and fMRI: all the while also addressing the challenge of missing modality data.

This work has certain limitations. Since this longitudinal study is ongoing, the experiments in this work are only carried out on a limited number of subjects, for whom the necessary two year follow-up data has been obtained [13]. EpiBioS4Rx has an expected enrollment of 300 subjects at completion. This will substantially increase the sample size, so we can test the robustness of the proposed methods, as well as investigate some of the applicable contemporary techniques. One such example is EmbraceNet [50], which has shown promise for multimodal classification with missing data, but needs more samples to properly train its deep network with more substantially more parameters, compared to ours.

V. CONCLUSION

In this work, we collected multicenter TBI subject data and investigated them for a binary late seizure classification task, using three modalities of interest (fMRI, dMRI, EEG). To handle missing data in certain subjects, we extended the NB classifier (following our hypothesized graphical model) to marginalize the missing modality, as well as implement standard imputation techniques, to be used with a chosen set of existing DR techniques and ML classifiers. In terms of AUC, while the NB estimator performed better than the imputation based individual modality base learners, the proposed IDSF algorithm performed the best within our experimental framework. IDSF attempted to capture both the shared and unique information, present within the multiple input modalities. To estimate the unique information in each modality, we also proposed RECC, which filtered already selected information based on correlation. Finally, following statistical analyses of the most frequently selected features, we found evidence that the fMRI alterations of the inferior temporal gyrus might serve as a potential biomarker of LPTS, and ultimately, PTE.

REFERENCES

[1] J. J. Gugger and R. Diaz-Arrastia, "Early Posttraumatic Seizures—Putting Things in Perspective," *JAMA Neurology*, vol. 79, no. 4, pp. 325–326, 04 2022.

[2] L. Piccenna, G. Shears, and T. J. O'Brien, "Management of post-traumatic epilepsy: An evidence review over the last 5 years and future directions," *Epilepsia open*, vol. 2, no. 2, pp. 123–144, 2017.

[3] D. J. Sharp, C. F. Beckmann, R. Greenwood, K. M. Kinnunen, V. Bonnelle, X. De Boissezon, J. H. Powell, S. J. Counsell, M. C. Patel, and R. Leech, "Default mode network functional and structural connectivity after traumatic brain injury," *Brain*, vol. 134, no. 8, pp. 2233–2247, 2011.

[4] "The epilepsy bioinformatics study for anti-epileptogenic therapy (EpiBioS4Rx) clinical biomarker: Study design and protocol," *Neurobiology of Disease*, vol. 123, pp. 110–114, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0969996118303115

[5] D. H. Wolpert, "Stacked generalization," *Neural networks*, vol. 5, no. 2, pp. 241–259, 1992.

[6] S. Džeroski and B. Ženko, "Is combining classifiers with stacking better than selecting the best one?" *Machine learning*, vol. 54, no. 3, pp. 255–273, 2004.

[7] G. Brown, "A new perspective for information theoretic feature selection," in *Artificial intelligence and statistics*. PMLR, 2009, pp. 49–56.

[8] Y.-D. Zhang, Z. Dong, S.-H. Wang, X. Yu, X. Yao, Q. Zhou, H. Hu, M. Li, C. Jiménez-Mesa, J. Ramirez *et al.*, "Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation," *Information Fusion*, vol. 64, pp. 149–187, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.22.22281402; this version posted October 25, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[9] J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. R. Bull, and N. Canagarajah, "Pixel-and region-based image fusion with complex wavelets," *Information fusion*, vol. 8, no. 2, pp. 119–130, 2007.

[10] J. Scheffer, "Dealing with missing data," 2002.

[11] J. Liu, X. Liu, Y. Zhang, P. Zhang, W. Tu, S. Wang, S. Zhou, W. Liang, S. Wang, and Y. Yang, "Self-representation subspace clustering for incomplete multi-view data," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 2726–2734.

[12] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[13] N. R. Temkin, "Preventing and treating posttraumatic seizures: the human experience," *Epilepsia*, vol. 50, pp. 10–13, 2009.

[14] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "Fsl," *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.

[15] R. K. Gupta, S. Saksena, A. Agarwal, K. M. Hasan, M. Husain, V. Gupta, and P. A. Narayana, "Diffusion tensor imaging in late posttraumatic epilepsy," *Epilepsia*, vol. 46, no. 9, pp. 1465–1471, 2005.

[16] M. N. Akbar, S. Ruf, M. La Rocca, R. Garner, G. Barisano, R. Cua, P. Vespa, D. Erdoğmuş, and D. Duncan, "Lesion normalization and supervised learning in post-traumatic seizure classification with diffusion mri," in *Computational Diffusion MRI*. Cham: Springer International Publishing, 2021, pp. 133–143.

[17] S. M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. E. Nichols, C. E. Mackay, K. E. Watkins, O. Ciccarelli, M. Z. Cader, P. M. Matthews *et al.*, "Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data," *Neuroimage*, vol. 31, no. 4, pp. 1487–1505, 2006.

[18] J. L. Stein, S. E. Medland, A. A. Vasquez, D. P. Hibar, R. E. Senstad, A. M. Winkler, R. Toro, K. Appel, R. Bartecek, Ø. Bergmann *et al.*, "Identification of common variants associated with human hippocampal and intracranial volumes," *Nature genetics*, vol. 44, no. 5, pp. 552–561, 2012.

[19] P. Vespa, M. Tubi, J. Claassen, M. Buitrago-Blanco, D. McArthur, A. G. Velazquez, B. Tu, M. Prins, and M. Nuwer, "Metabolic crisis occurs with seizures and periodic discharges after brain trauma," *Annals of Neurology*, vol. 79, pp. 579–590, 2016.

[20] R. Faghihpirayesh, S. Ruf, M. La Rocca, R. Garner, P. Vespa, D. Erdogmus, and D. Duncan, "Automatic detection of eeg epileptiform abnormalities in traumatic brain injury using deep learning," in *IEEE Engineering in Medicine and Biology Society*, 2021.

[21] J. A. Kim, E. J. Boyle, A. C. Wu, A. J. Cole, K. J. Staley, S. Zafar, S. S. Cash, and M. B. Westover, "Epileptiform activity in traumatic brain injury predicts post-traumatic epilepsy," *Annals of Neurology*, vol. 83, pp. 858–862, 2018.

[22] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[23] E. F. Chang, M. Meeker, and M. C. Holland, "Acute traumatic intraparenchymal hemorrhage: risk factors for progression in the early post-injury period," *Neurosurgery*, vol. 58, no. 4, pp. 647–656, 2006.

[24] J. Ashburner and K. J. Friston, "Unified segmentation," *NeuroImage*, vol. 26, no. 3, pp. 839–851, Jul. 2005.

[25] S. Whitfield-Gabrieli and A. Nieto-Castanon, "Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks," *Brain connectivity*, vol. 2, no. 3, pp. 125–141, 2012.

[26] J. Crinion, J. Ashburner, A. Leff, M. Brett, C. Price, and K. Friston, "Spatial normalization of lesioned brains: performance evaluation and impact on fmri analyses," *Neuroimage*, vol. 37, no. 3, pp. 866–875, 2007.

[27] S. M. Andersen, S. Z. Rapcsak, and P. M. Beeson, "Cost Function Masking during Normalization of Brains with Focal Lesions: Still a Necessity?" *NeuroImage*, vol. 53, no. 1, pp. 78–84, Oct. 2010.

[28] J. Ashburner, "Computational anatomy with the spm software," *Magnetic resonance imaging*, vol. 27, no. 8, pp. 1163–1174, 2009.

[29] A. V. Faria, S. E. Joel, Y. Zhang, K. Oishi, P. C. van Zjil, M. I. Miller, J. J. Pekar, and S. Mori, "Atlas-based analysis of resting-state functional connectivity: Evaluation for reproducibility and multi-modal anatomy–function correlation studies," *Neuroimage*, vol. 61, no. 3, pp. 613–621, 2012.

[30] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman, "Missing value estimation methods for dna microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, 2001.

[31] D. Erdogmus, U. Ozertem, and T. Lan, *Information Theoretic Feature Selection and Projection*, 12 2007, vol. 83, pp. 1–22.

[32] H. Hotelling, "Relations between two sets of variates," in *Breakthroughs in statistics*. Springer, 1992, pp. 162–190.

[33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[34] N.-F. Jie, M.-H. Zhu, X.-Y. Ma, E. A. Osuch, M. Wammes, J. Théberge, H.-D. Li, Y. Zhang, T.-Z. Jiang, J. Sui *et al.*, "Discriminating bipolar disorder from major depression based on svm-foba: efficient feature selection with multimodal brain imaging data," *IEEE transactions on autonomous mental development*, vol. 7, no. 4, pp. 320–331, 2015.

[35] M. N. Akbar, M. La Rocca, R. Garner, D. Duncan, and D. Erdoğmuş, "Prediction of epilepsy development in traumatic brain injury patients from diffusion weighted mri," in *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, ser. PETRA '20, New York, NY, USA, 2020.

[36] J. Chapman and H.-T. Wang, "Cca-zoo: A collection of regularized, deep learning based, kernel, and probabilistic cca methods in a scikit-learn style framework," *Journal of Open Source Software*, vol. 6, no. 68, p. 3823, 2021. [Online]. Available: https://doi.org/10.21105/joss.03823

[37] O. Özdenizci and D. Erdoğmuş, "Stochastic mutual information gradient estimation for dimensionality reduction networks," *Information Sciences*, vol. 570, pp. 298–305, 2021.

[38] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve." *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[39] K. Hajian-Tilaki, "Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation," *Caspian journal of internal medicine*, vol. 4, no. 2, p. 627, 2013.

[40] A. Tenenhaus and M. Tenenhaus, "Regularized generalized canonical correlation analysis," *Psychometrika*, vol. 76, no. 2, pp. 257–284, 2011.

[41] P. L. Williams and R. D. Beer, "Nonnegative decomposition of multivariate information," *arXiv preprint arXiv:1004.2515*, 2010.

[42] J. Sourati, D. Erdogmus, M. Akcakaya, S. C. Kazmierczak, and T. K. Leen, "A NOVEL DELTA CHECK METHOD FOR DETECTING LABORATORY ERRORS Northeastern University University of Pittsburgh Oregon Health & Science University Portland , OR , USA National Science Foundation," pp. 0–5, 2015.

[43] O. J. Dunn, "Multiple comparisons among means," *Journal of the American statistical association*, vol. 56, no. 293, pp. 52–64, 1961.

[44] E. S. Lutkenhoff, V. Shrestha, J. R. Tejeda, C. Real, D. L. McArthur, D. Duncan, M. La Rocca, R. Garner, A. W. Toga, P. M. Vespa *et al.*, "Early brain biomarkers of post-traumatic seizures: initial report of the multicentre epilepsy bioinformatics study for antiepileptogenic therapy (epibios4rx) prospective study," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 91, no. 11, pp. 1154–1157, 2020.

[45] R. Immonen, N. G. Harris, D. Wright, L. Johnston, E. Manninen, G. Smith, A. Paydar, C. Branch, and O. Grohn, "Imaging biomarkers of epileptogenecity after traumatic brain injury – preclinical frontiers," *Neurobiology of Disease*, vol. 123, pp. 75–85, 2019, antiepileptogenesis following Traumatic Brain Injury.

[46] A. Irimia, S.-Y. M. Goh, A. C. Wade, K. Patel, P. M. Vespa, and J. D. Van Horn, "Traumatic brain injury severity, neuropathophysiology, and clinical outcome: insights from multimodal neuroimaging," *Frontiers in neurology*, vol. 8, p. 530, 2017.

[47] A. Singh, C. Westlin, H. Eisenbarth, E. A. Reynolds Losin, J. R. Andrews-Hanna, T. D. Wager, A. B. Satpute, L. F. Barrett, D. H. Brooks, and D. Erdogmus, "Variation is the norm: Brain state dynamics evoked by emotional video clips," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2021, pp. 6003–6007.

[48] N. Memarian, S. Kim, S. Dewar, J. Engel, and R. J. Staba, "Multimodal data and machine learning for surgery outcome prediction in complicated cases of mesial temporal lobe epilepsy," *Computers in Biology and Medicine*, vol. 64, pp. 67–78, 2015.

[49] Y. Yang, N. D. Truong, J. K. Eshraghian, C. Maher, A. Nikpour, and O. Kavehei, "A multimodal ai system for out-of-distribution generalization of seizure identification," *IEEE Journal of Biomedical and Health Informatics*, 2022.

[50] J.-H. Choi and J.-S. Lee, "Embracenet: A robust deep learning architecture for multimodal classification," *Information Fusion*, vol. 51, pp. 259–270, 2019.