SUPPLEMENTAL MATERIAL

Expanded Materials & Methods:

Anchoring

After a subarachnoid bleed due to aneurysm, patients are closely monitored for signs of DCI onset from PBD 3 to PBD 14–21. This large window for presentation of disease makes direct comparison across patients over time difficult. Here we use the diagnosis of DCI as the temporal anchor to align the data so that we can identify a physiologic signal as the disease develops, despite varied times of onset. For DCI positive patients onset was determined as the earliest point that any of the definition criteria were met (DND or new focal deficit and/or new infarct due to VSP). Some patients with DCI also demonstrated arterial spasm on vessel imaging, again adjudicated by consensus, prior to clinical symptoms. In these cases the earliest indication of vasospasm was used as the anchor in order to reduce the effects of causality leakage [33]. For patients who did not develop DCI, PBD 7 was used as the anchor since it is the clinically-accepted peak time of DCI risk and when surveillance scans typically confirm the absence of angiographic vasospasm triggering de-escalation. Patients with only non-symptomatic angiographic VSP were not included in either category in the final analysis. A 14-day window of physiological data (±7 days from anchor), was extracted for analysis and comparison between the DCI+ and the DCI– control group.

Machine Learning

Ensemble classifiers are based on soft voting, which predicts the class labels based on the predicted probabilities for classifiers, i.e.:

$$\hat{y} = \arg\max_{i} \sum_{j=1}^{m} w_j p_{ij}$$

Where $\hat{y}$ is the computed probability by the voting classifier, w_j is the weight assigned to the j^{th} classifier and $i \in \{0,1\}$ for a binary classifier.

Nested Cross-Validation

Model selection using the same data to tune and evaluate models leads to over-fitting. Model selection with nested cross-validation uses a series of train/validation/test splits to avoid this problem. The inner loop (executed here with Grid search) tunes the parameters of the model (built on the training set) by maximizing the performance on the validation set. In the outer loop,

generalization error is estimated by the median of the performance scores on the testing set over several (in this case, 5) dataset splits.

Figure I: Features over time

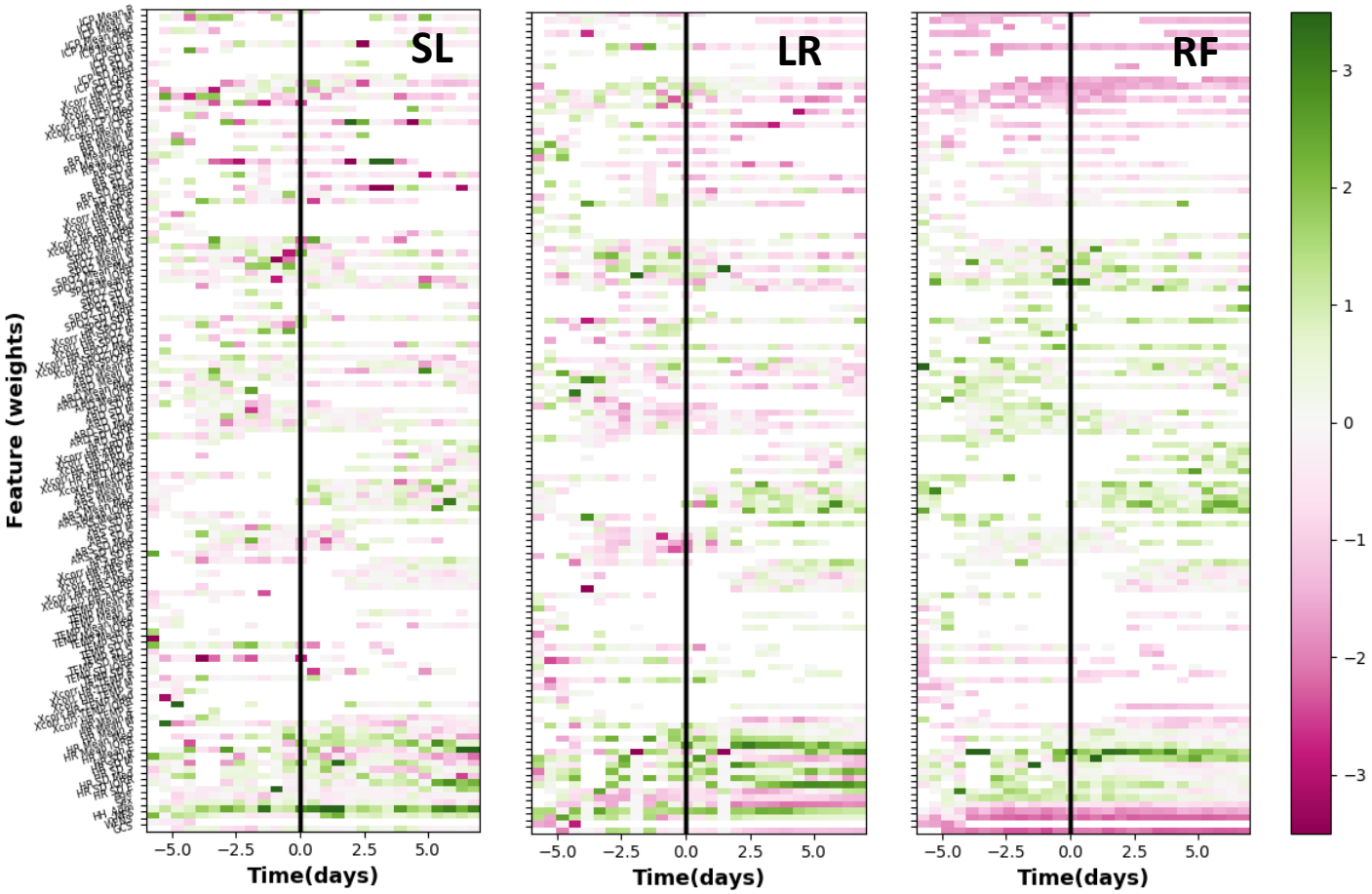

Figure II: Performance on Angiographic Vasospasm dataset.