Fusion Approaches to Predict Post-stroke Aphasia Severity from Multimodal Neuroimaging Data

Saurav Chennuri, Sha Lai, Anne Billot, Maria Varkanitsa, Emily J. Braun, Swathi Kiran, Archana Venkataraman, Janusz Konrad, Prakash Ishwar and Margrit Betke

Workshop on Computer Vision for Automated Medical Diagnosis, October 2, 2023

Task Definition

Aphasia:
- Language disorder characterized by impairment in language comprehension & speech production
- Affects ~1/3 of stroke survivors

Predicting aphasia evolution is important:
- Clinicians □ for choosing treatment
- Patients □ to engage in recovery process

Our task:
- Develop regressor that predicts patients' aphasia severity as measured by the Western Aphasia Battery (WAB) test score
- Using multimodal neuroimaging data.

Challenges & Contributions

Challenges:
- Imbalance between large number of features & limited number of subjects
- Danger of model overfitting

Previous work:
- Oracle feature selection guided by *all* data
- Danger of "data leakage" from training to testing datasets, i.e., questionable generalizability

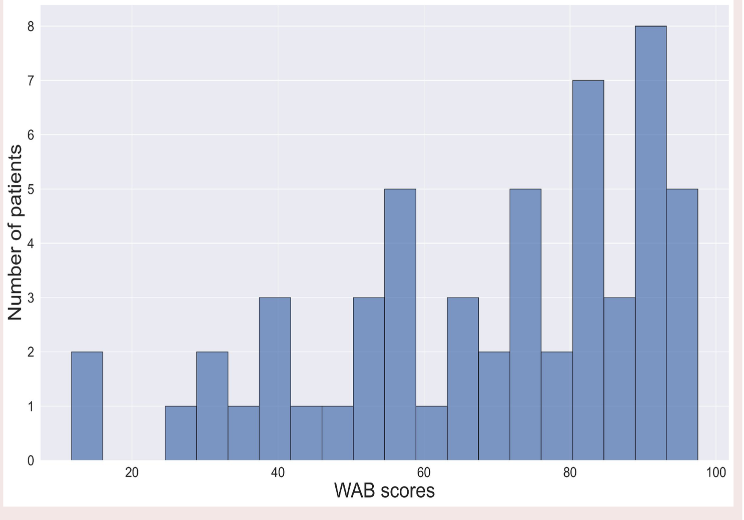
Contributions:
- Supervised Feature Selection Algorithm
- Two fusion strategies applied to aphasia data
- Determining predictive importance of different multimodal combinations of feature sets

Previous Work

Fusion approaches:
- Early Fusion (feature fusion) and Late Fusion (prediction fusion) [Huang et al., 2020, Holste et al., 2021]
- Binary classifiers, predicting response to treatment (changes in treatment outcome): Random Forests (RFs), Support Vector Machines (SVMs) [papers by our team, 2020, 2021, 2022]
- Regression with oracle feature selection using
  - stacked multimodal predictions [Pustina et al., 2017]
  - stacked multimodal features, [Kristinsson et al., 2021]

Data

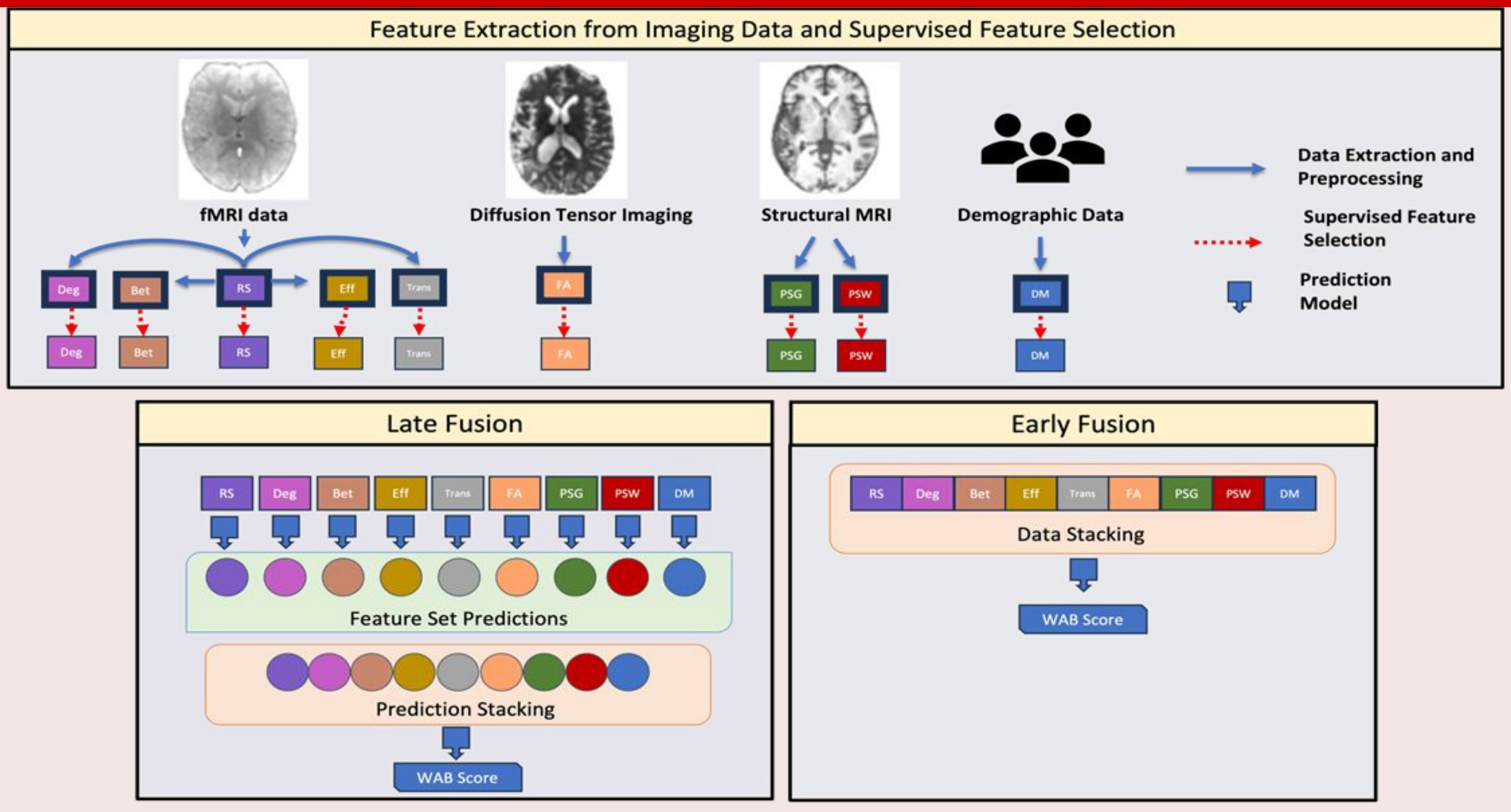
- **55 patients**
- **Imaged at:**
  1. Boston U.
  2. Johns Hopkins U.
  3. Northwestern U.



- **9 feature sets**
- **945 dimensions**

Data	Feature Set	Symbol	#
Dem.	Age, education, time poststroke	DM	3
DTI	Fractional anisotropy	FA	12
fMRI	Bidirectional correlations	RS	625
	Betweenness	Bet	50
	Degree	Deg	50
	Efficiency	Eff	50
	Transitivity	Trans	50
MRI	Percent of spared gray matter	PSG	69
MRI	Percent of spared white matter	PSW	36

Method



Supervised Feature Selection Algorithm

- **Task:** Determine how many & which features per feature modality
- **Inspired by:** nested cross-validation *[[train:validate]:test]*
- **Each split of *[train:validate]* used for selecting:**
  - number of features in each feature set
  - specific features within feature set
  - hyperparameters of SVR or RF predictors
- **Feature sorting based on:** empirical cross-correlation values between features & WAB scores <u>only on train data</u>

Methodology (key steps)
- For value of k under consideration: features occurring among the top k features most often across all training sub-folds are chosen to train SVR or RF to predict WAB scores using the entire [train] fold & tune its hyperparameters using the same sub-folds created previously
- Among the resulting models M_k, cross-validated with Root Mean Square Error (RMSE), model M_{k^*} with smallest RMSE is chosen.
- Note:
  - Optimal features are specific to [train] set
  - Different features may be selected on each [test] set & modality

Cross-Validation Experiments & Results

1. **Same *[train:validate]:test]* sets as for feature selection:** *[[45:5]:5]* patients, 10 inner folds, 11 outer folds.

2. Feature reduction:

	DM	FA	PSG	PSW	RS	Bet	Deg	Eff	Tran
SVR	2	6	18	12	75	42	13	7	12
RF	2	5	7	31	116	33	37	20	14

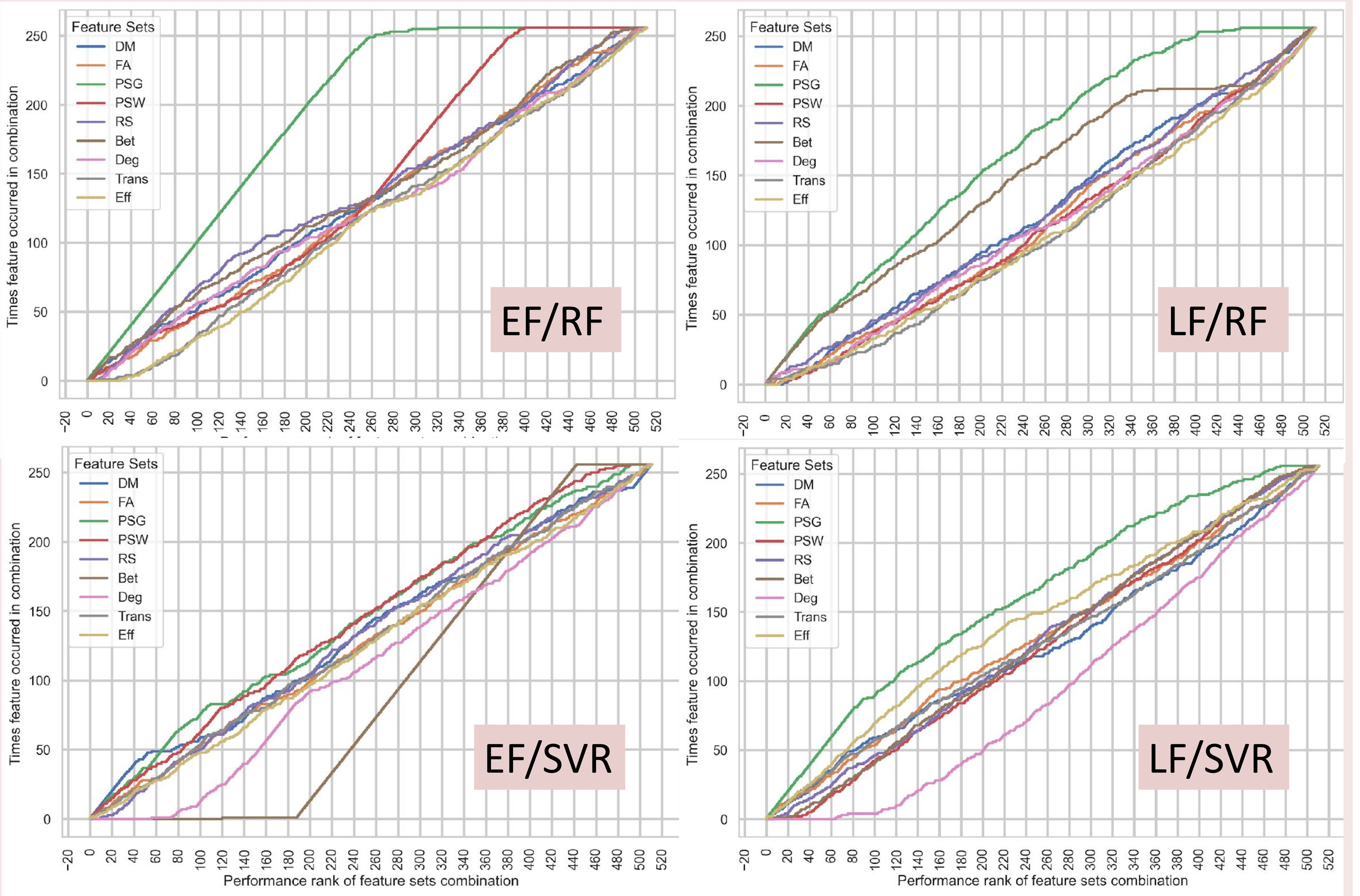
3. Worse performance when using:
- all features
- single features

4. Overall Performance:

	Early Fusion		Late Fusion	
	RF	**SVR**	**RF**	**SVR**
	Supervised Feature Selection			
Mean RMSE	17.41	16.72	17.45	16.85
Mean r	0.48	0.64	0.58	0.58
	Oracle Selection			
Mean RMSE	15.19	14.4	13.78	13.36
Mean r	0.80	0.77	0.80	0.81

Feature Set Importance

- Feature set cumulative occurrence plots:
  - Number of times each feature set appears within top-ranked combinations
  - Rank based on mean-fold RMSE metric
  - Plots are for rank values 1 through 511
- **Most import feature sets:**
  1. PSG = Percent spared gray matter
  2. PSW = Percent spared white matter
  3. DM = Demographics



Discussion & Conclusions

- **Early Fusion (EF) versus Late Fusion (LF):**
  - Best mean-fold RMSE for RF & SVR lower for EF than LF with Supervised Feature Selection
  - LF better than EF for Oracle Selection
  - But std. dev. of RMSE is high
  - Test of statistical significance among top-feature set combinations □ No clear winner among EF & LF approaches
- Feature reduction + multimodal feature combinations:
  - Useful practical strategy
  - Improves overall interpretability of results
- Effective predictors of aphasia severity: Feature combinations with Percent Spared Gray Matter