Northeastern University
EECE5644 – Pattern Recognition and Machine Learning
April 2018, Asha Chen-Phang, Emily Dutile, Nate Otenti, Tristan Sweeney

Classification in High-Resolution Brain Scans Using Apache Spark: Project Work Distribution

Asha Chen-Phang:

- Data analysis in Jupyter Notebook (python)
- Aided with parameter tuning within the model code of the Spark/Scala program
- Prepared, wrote and edited the Report with images and references
- Prepared the Presentation
- Wrote the Readme with instructions on how to build and run locally as well as run on EMR

Emily Dutile:

- Implemented and configured Makefile needed for standalone and AWS runs for Apache Spark
- Executed AWS/EMR Runs and gathered results for analysis
- Implemented Random Forest Model Training in scala
- Implemented Random Forest Prediction Phase in scala
- Contributed to Feature Extraction and preprocessing in Jupyter Notebook
- Report contributions

Nate Otenti:

- Model Comparison Tests in python (Jupyter notebooks)
- Apache Maven integration - pom file
- Contributed to Feature Extraction and preprocessing in Jupyter Notebook
- Presentation contributions
- Report contributions

Tristan Sweeney:

- Random Forest in python (Jupyter Notebook)
- KNN in python (Jupyter Notebook)
- SVM in python (Jupyter Notebook)
- Executed numerous runs with entire dataset and pre-processed dataset for analysis
- SVM in python (Jupyter Notebook)
- Report contributions