Report on the study of Joseph Son, Mattheus Bezerra, and D. Ben Knoble

Over the course of the Spring 2019 semester, the research group headed by Martin Styner and Mahmoud Mostapha and consisting of undergraduates Joseph Son, Mattheus Bezerra, and D. Ben Knoble explored Machine and Deep Learning, with particular emphasis on applications to MRI features of young children.

The beginning of the semester covered a fast-paced introduction to the fields of Machine Learning and Data Science. The researchers developed insight about the generalization of standard 2D statistical analysis and regression to higher-dimensional spaces, learned about the gradient descent technique used to solve high-order non-linear equations, and discussed the inherent risks of stochastic methods. The team further spent time with Google's ML Crash Course and other web resources in order to get a grasp on the fundamentals of the problem space [Goo]. The undergraduates finally dug into the idea of Deep Learning and Neural Networks, the layered approach to creating these analytical systems.

As the semester continued, the team moved into consideration of the particular datasets it would be working with. This included visualizing the features of processed MRIs, broken down into Regions of Interest (ROIs), in addition to exploring the demographics of the sample. Another two data sets were also provided in BASC and BRIEF test scores for the 6 year old subjects. Their behavioral temperaments were recorded to determine or verify that a child may have an abnormality. The team was interested in the assessing data relevant both to conte and twin subjects. Researchers also attempted to formulate different classifications of subjects from a biological, psychological, and behavioral background, in order to develop a model on which to train eventual classification networks.

The team generated ADHD risk groups by creating regions based on t-scores of certain features from BASC and BRIEF results that are known to be highly correlated with ADHD: Attention Problems, HyperActivity, Inhibition, and Working Memory. After each subject was assigned a risk label, we projected subjects with available cortical thickness and surface area data with a manifold learning technique known as UMAP. The risk group boundaries and UMAP projections are listed below.

There are several different limitations of the data, including the small sample sizes and the lack of certain data points for individual subjects. It was found that there would be far too many features to be learned via deep learning for the given amount of subjects in the study. The data samples taken cannot be considered a real world projection due to its relatively low and specific sample size. There is also a much higher ratio of low-risk subjects which does not allow for an easily constructed classification network. A lookup table was also created due to the differences in naming conventions between data sheets so as to understand the discrepancies and missing subjects that may have had their brain features measured, taken the BASC/BRIEF, or a combination of the two. Techniques under consideration to handle these issues include oversampling, data interpolation, and feature-set reduction via dimensionality reduction techniques such as auto-encoders. Support vector machines are also being considered due to their ability to separate and classify the given data at a lower dimensional level.

ADHD Risk Groups

High Risk (HR)

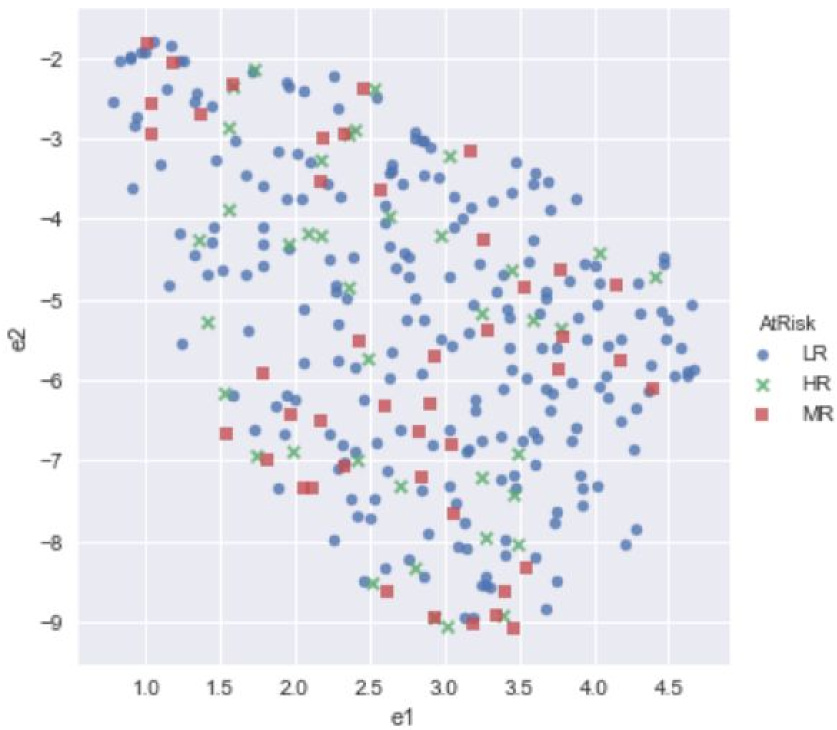
1. BASC: AttProb >= 65 || HyperActivity >= 65
2. BRIEF: Inhibit >= 65 || (Inhibit < 65 & Working Memory >= 65)

Moderate Risk (MR)

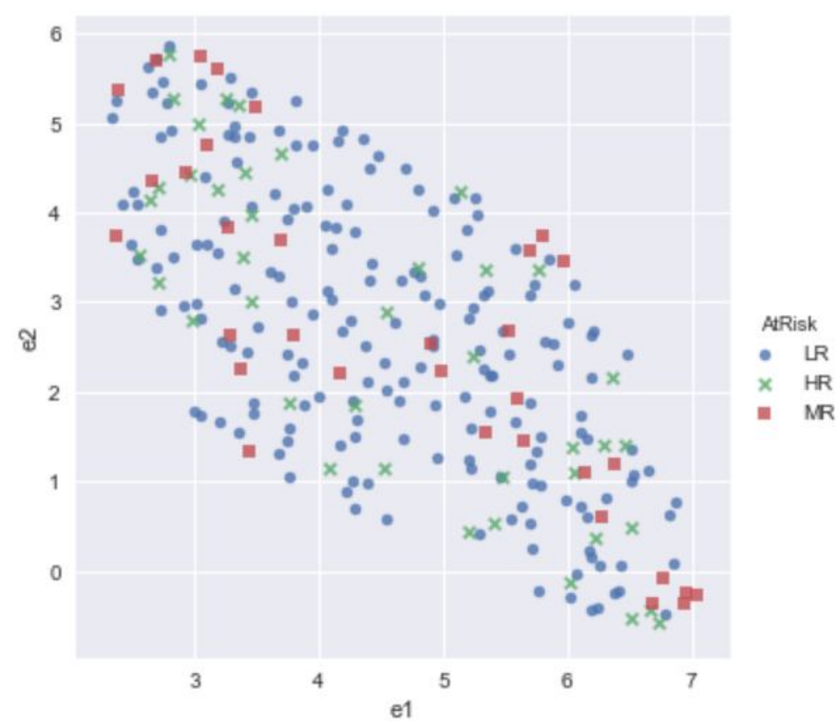
1. BASC: AttProb < 65 & HyperActivity < 65
2. BRIEF: Inhibit >= 65 || (Inhibit < 65 & Working Memory >= 65)
   OR
1. BASC: AttProb >= 65 & HyperActivity >= 65
2. BRIEF: Inhibit < 65 & Working Memory < 65

Low Risk (LR)

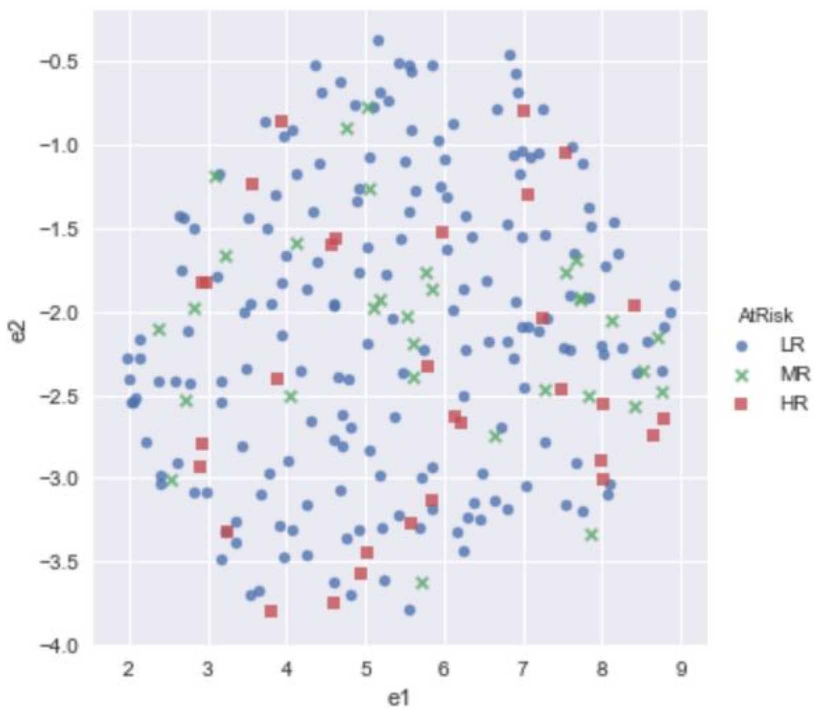
1. BASC: AttProb < 65 & HyperActivity < 65
2. BRIEF: Inhibit < 65 & Working Memory < 65



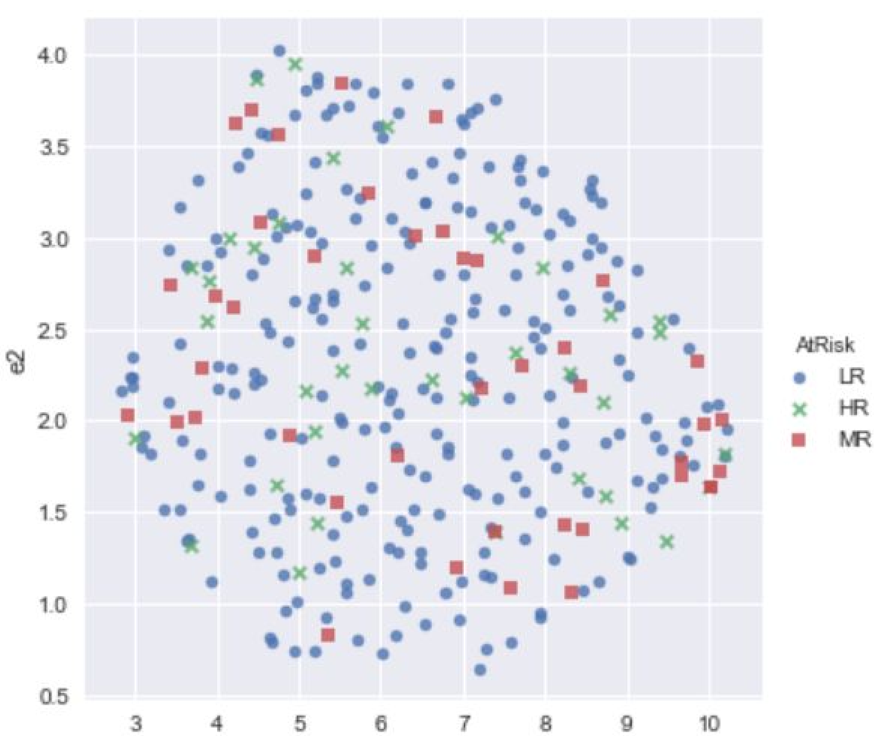
UMAP Projection of SA data at 1 year



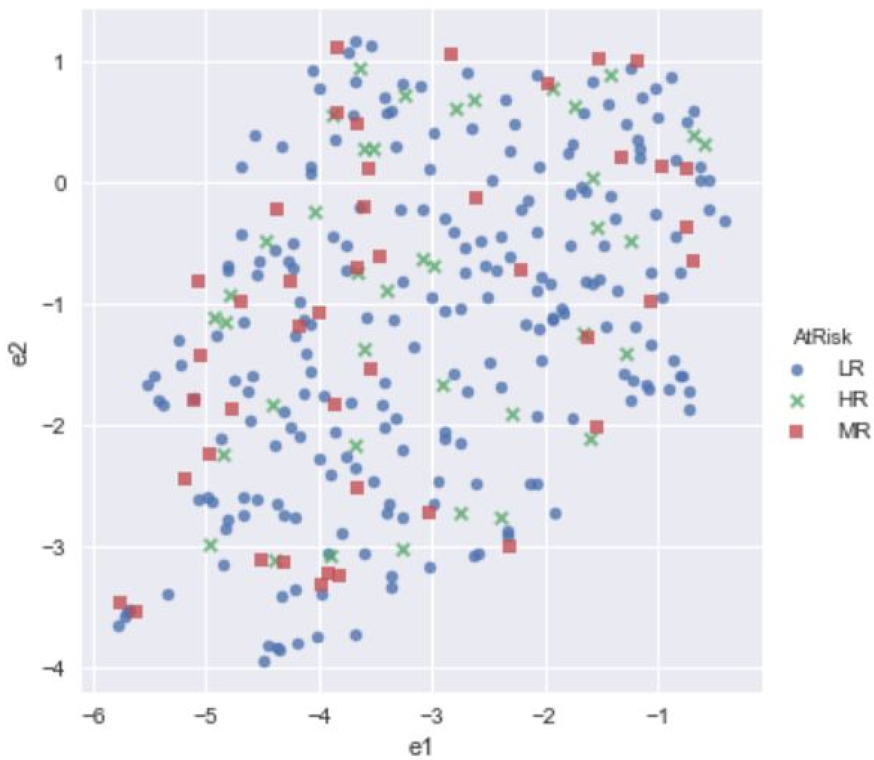
UMAP Projection of SA data at 2 years



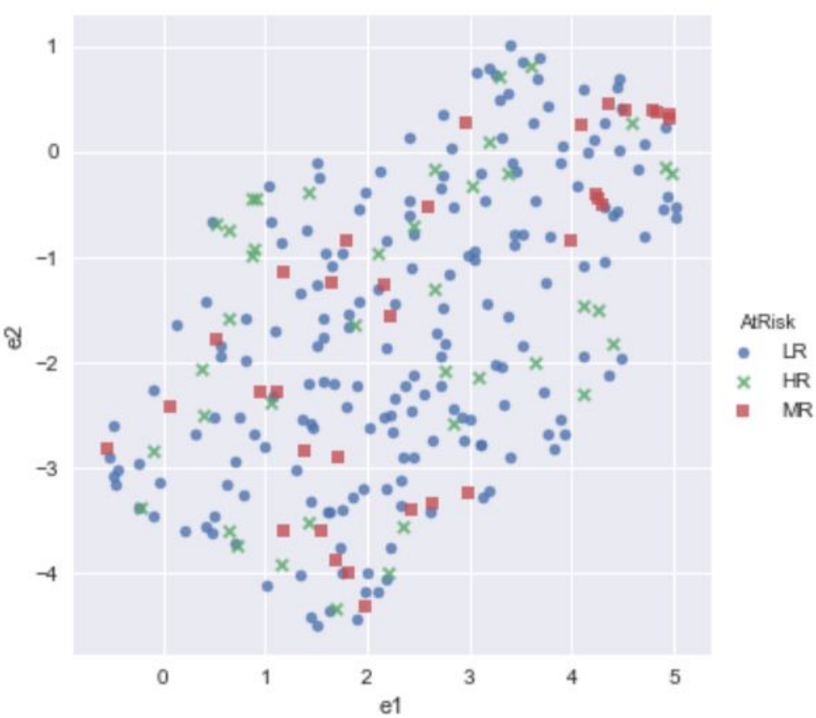
UMAP Projection of SA data at 4 years



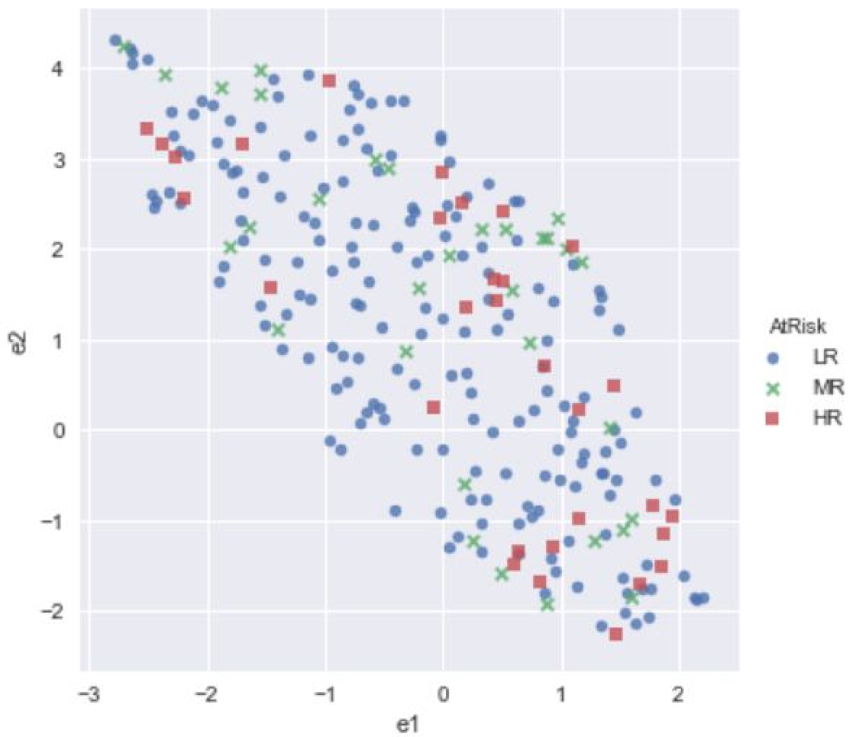
UMAP Projection of SA data at 6 years



UMAP Projection of CT data at 1 year



UMAP Projection of CT data at 2 years



UMAP Projection of CT data at 4 years

UMAP Projection of CT data at 6 years

Lookup Table Findings

Study	BASC and MRI Scan	BASC Only	MRI Only
CONTE 1 Year	131	103	117
CONTE 2 Year	119	115	69
CONTE 4 Year	123	111	39
CONTE 6 Year	171	63	1
TWIN 1 Year	171	49	134
TWIN 2 Year	143	77	74
TWIN 4 Year	123	97	39
TWIN 6 Year	167	53	0

References

[BRRG18] David Berthelot, Colin Rael, Aurko Roy, and Ian J. Goodfellow. Understanding and improving interpolation in autoencoders via an adversarial regularizer. *CoRR*, abs/1807.07543, 2018.

[CCG+18] David Charte, Francisco Charte, Salvador García, Mara J. del Jesus, and Francisco Herrera. A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines. *Information Fusion*, 44:78—96, nov 2018.

[GKG18] John H. Gilmore, Rebecca C. Knickmeyer, and Wei Gao. Imaging structural and functional brain development in early childhood. *Nature Reviews Neuroscience*, 19(3):123—137, feb 2018.

[Goo] Google. Machine learning crash course. https://developers.google.com/machine-learning/crash-course/. Accessed on 2019-01-22.

[JRS05] Kelly Pizzitola Jarratt, Cynthia A. Riccio, and Becky M. Siekierski. Assessment of attention deficit hyperactivity disorder (ADHD) using the BASC and BRIEF. *Applied Neuropsychology*, 12(2):83—93, jun 2005.

[LBH15] Yann LeCun, Yoshua Bengio, and Georey Hinton. Deep learning. *Nature*, 521(7553):436—444, may 2015.

[Mor18] Majid Mortazavi. Umap benchmark. https://github.com/mmortazavi/UMAP_Nonlinear-Dimensionality-Reduction_Benchmark/blob/master/UMAP_Benchmark.ipynb, May 2018. Accessed on 2019-03-08.

[Mos18] Mahmoud Mostapha. Comp 562: Introduction to machine learning. http://comp562fall18.web.unc.edu, Aug 2018. Accessed on 2019-01-10.

[Say18] Sayantini. edureka: Autoencoders tutorial : A beginner's guide to autoencoders. https://www.edureka.co/blog/autoencoders-tutorial/, Oct 2018. Accessed on 2019-02-13.

[SHS+18] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, and Demis Hassabis. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140—1144, dec 2018.

[YLZ+ 18] Qianye Yang, Nannan Li, Zixu Zhao, Xingyu Fan, Eric I-Chao Chang, and Yan Xu. MRI image-to-image translation for cross-modality image registration and segmentation. *CoRR*, abs/1801.06940, 2018.