

Iowa Initiative for Artificial Intelligence

Final Report

Project title:	Predict Huntington’s Disease (HD) using structural and functional neuroimaging data		
Principal Investigator:	Jordan Schultz		
Prepared by (IIAI):	Yanan Liu		
Other investigators:			
Date:			
Were specific aims fulfilled:		Y	
Readiness for extramural proposal?		Y	
If yes ... Planned submission date			
Funding agency			
Grant mechanism			
If no ... Why not? What went wrong?			

Brief summary of accomplished results:

We have developed and validated a Random Forest model to accurately predict Huntington's Disease (HD) using structural and functional neuroimaging data and demonstrated excellent performance. The prediction accuracy is 99% and F1 is 0.99.

Research report:

Aims (provided by PI):

Aim 1: Assess the accuracy of a machine-learning classification algorithm to identify HD from non-HD participants using neuroimaging data.

Aim 2: Identify regions-of-interest impacted by altered neurodevelopment in HD.

Data:

The Kids-HD dataset currently includes over 300 total participants, including controls and kids with HD. Many of those participants have multiple neuroimaging studies available. Specifically, there are nearly 500 neuroimaging studies available for analysis. Baseline demographics of these participants are included in Table 1. Participants in the Kids-HD study have also undergone extensive clinical testing. Previous studies have demonstrated that there may be cognitive differences amongst participants based on CAG repeat length. Cognitive tests include a timed pegboard test, the Wechsler Intelligence Scale for Children, and the DelisKaplan Executive Function System, amongst others. Overall, 531 neuroimaging studies were used for prediction.

Table 1 Baseline demographics

	HD	Controls	p-value
N (total visits)	83 (136)	235 (343)	NA
Age, mean (SD)	13.6 (4.2)	12.7 (3.8)	0.060
Female, N (%)	51 (61.4)	122 (51.9)	0.171
CAG, mean (SD)	45.6 (5.6)	20.2 (4.0)	<0.001
Predicted Years from Onset, mean (SD)	44.7 (14.2)	NA	NA

AI/ML Approach:

In this study, supervised machine learning algorithm was implemented for prediction using Python. As many extracted features may be noisy, or highly correlated with each other, Random Forest (RF) algorithm was selected to predict tumor/nevus group and performance assessed using 5-fold-cross-validation (Figure 1). Accuracy and F1-score were calculated to compare performance of different tests.

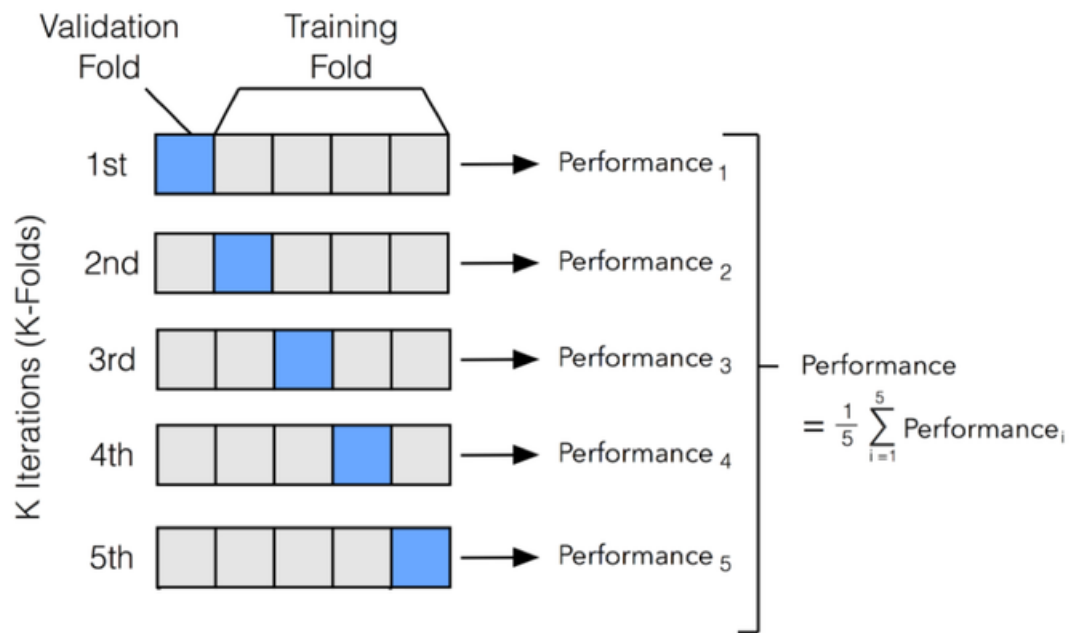


Figure 1. Representation of 5-fold cross-validation technique

Results:

We have developed and validated a Random Forest model to accurately predict Huntington's Disease (HD) using structural and functional neuroimaging data. The prediction accuracy is 99% and F1 is 0.99.

Table 2 Confusion matrix

	Predicted Control	Predicted HD
Control	341	2
HD	1	187

Top 10 features from Random Forest model are listed in descending order: yto, amo, cag, put, bg, pall.l, put.r, acc.r, caud, bg.r.

Table 3 Feature importances of top 10 features

Feature name	Feature Importances
yto	0.168
amo	0.145
cag	0.142
put	0.015
Bg	0.014
Pall.l	0.014
Put.r	0.013
Acc.r	0.012
Caud	0.011
Bg.r	0.010
