# Fine-grained prediction of Huntington's disease progression using a stacked ensemble approach

**Maitrei Kohli[1], Dorian Pustina[3], John Warner[3], Rachael I. Scahill[2], Sarah J. Tabrizi[2], Daniel C. Alexander[1], Peter A. Wijeratne[1]**

**1: Centre of Medical Image Computing, Department of Computer Science, UCL; 2 : Huntington's Disease Centre, Institute of Neurology, UCL**
**3: CHDI Foundation, Inc.**

**We applied a *stacked ensemble* based approach for predictive-classification of 184 HD gene positive subjects from TRACK-HD dataset. Our model sets a new benchmark for fine-grained classification of HD, & thus sets the ground for clinically useful future work such as prediction of onset**

## 1. Background

- Classification of individual disease state in Huntington's disease (HD) is key for stratification of patients in clinical trials
- However, there has been no coordinated effort to assess benefits of using Machine learning (ML) based Stacking ensemble approach for making patient-specific & data-driven predictions of disease state in HD

## 2. Objectives

- We developed a novel ML framework that facilitates predictive-classification of individual-level HD disease state; the framework classifies subjects according to:

(a) binary HD disease state (i.e., Pre-manifest HD (PreHD); manifest HD)

(b) fine-grained disease state (i.e., PreHD A; PreHD B; HD1; HD2)

## 3. Methods

- ML framework comprises of:
  (a) Tier 1 (Base models): array of 6* standard ML models; trained & fit on actual training-data using repeated 10-fold cross-validation
  (b) Tier 2 (meta-model): trained on predictions made by Tier 1 models; learns how to best combine these predictions to reduce variance & generalisation error
- **Data** [+]: baseline cross-sectional data from 184 HD gene-positive participants from the TRACK-HD dataset, which includes clinical, imaging & genetic data

**Imaging** → Caudate; Pallidum; Putamen; .....
**Genetic** → CAG Repeat Length
**Clinical** → SDMT; SWRT; TMS; TFC
**General** → Age; Sex

**Figure 1: Input Features used for classification**

- **Subjects**: Details about how subjects were grouped in various classes can be found at - Tabrizi et al., 2013**

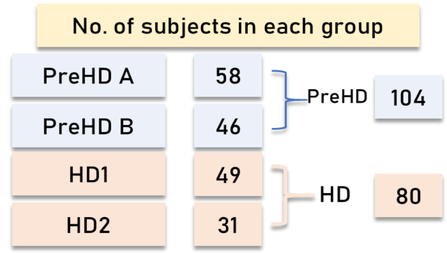


**Figure 2: No. of subjects in each class/group**

***ML models used**
**Base-models** → lr = logistic regression; knn = k nearest neighbour; cart = decision trees; svm = support vector machine; bayes = Gaussian Naïve Bayes; MLP = multi-layer perceptron
**Meta-model (Stacking)** → Gaussian Naïve Bayes
**Comparative ensemble methods** → ; AdaB = adaboost; XGB = extreme gradient boosting; RF = random forest

## 4. Results

### Binary classification (PreHD vs HD)

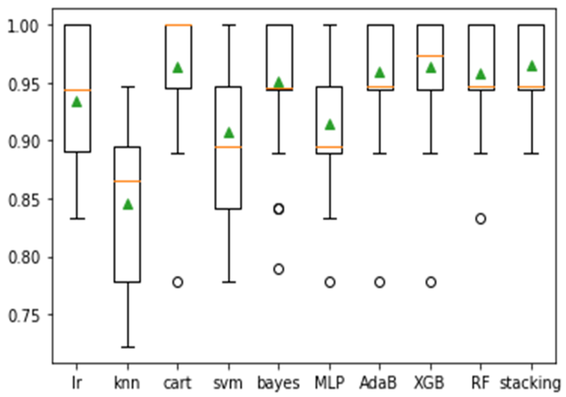


**Figure 3: Binary classification accuracy**

- Stacked ensemble model achieved an accuracy of 96.5%±3.6 in binary-classification task
- It also performed better than base-models & comparative ensemble models albeit non-significantly

### Fine-grained classification (PreHD A; PreHD B; HD1; HD2)

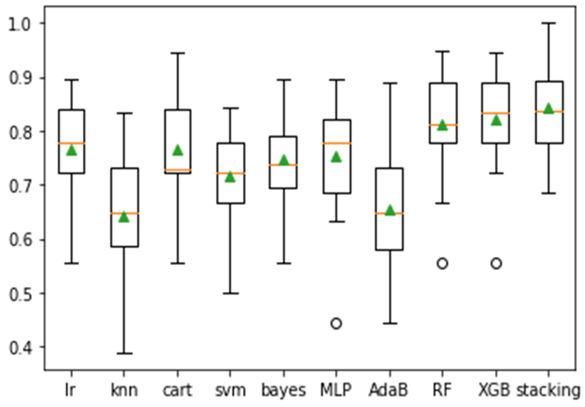


**Figure 4: Fine-grained classification accuracy**

- Stacked model achieved an accuracy of 85.2% ±8.0 in fine-grained classification task
- It also performed significantly (>9%) better than base-models & comparative ensemble models

*Our Stacked ensemble approach sets a new benchmark for fine-grained classification task*

## 5. Conclusion

- Our stacked model achieved best predictive-classification fine-grained classification tasks
- These encouraging results indicate that the Stacked ensemble approach might potentially be a powerful tool for making clinically useful predictions such as time-to-onset → Future work

**[+]Data: Full list of Imaging input Features used:**
(1.) Caudate; (2.) Pallidum; (3.) Putamen; (4.) Accumbens area; (5.) Lateral vents; (6.) Thalamus proper; (7.) Temporal; (8.) Frontal; (9.) Occipital; (10.) Parietal; (11.) Sensory Motor; (12.) Insula; (13.) Cingulate; (14.) Insula white matter

**[+]Full forms of clinical input Features used:**
SDMT → symbol digit modality test; SWRT → stroop word reading test; TMS → total motor score; TFC → total functional capacity

******Tabrizi, Sarah J., et al. "Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data." *The Lancet Neurology* 12.7 (2013): 637-649.

maitrei.kohli@ucl.ac.uk