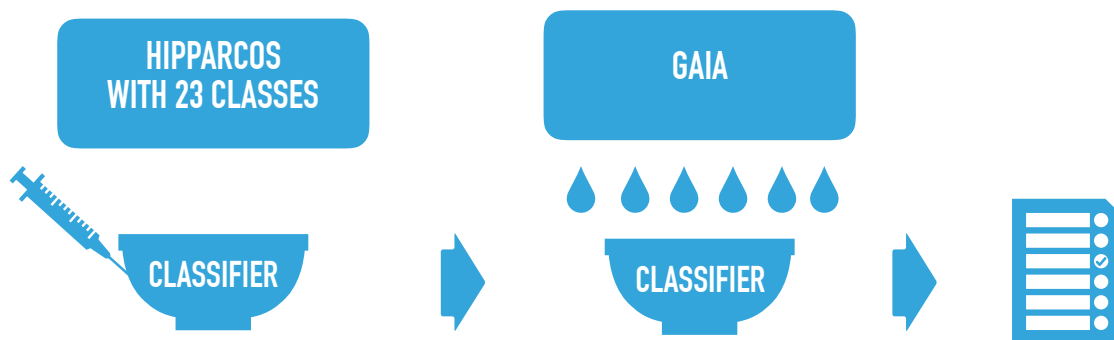# FINITE DIRICHLET MIXTURE MODELS FOR CLASSIFICATION AND DETECTION OF NEW CLASSES OF VARIABLE STARS

Prince John
Dr. Maria Süveges
Prof. Alessandra R Brazzale

GOALS

- To implement a statistical model which can classify the awaited GAIA dataset into these 23 classes.

- To implement a statistical model which can detect new classes of variable stars in the GAIA dataset.

GOAL 1 : CLASSIFICATION MODEL

Fit each of the 23 classes of variable stars to a finite mixture of Dirichlet distributions and then forming an ensemble of all the 23 mixtures

For a D dimensional dataset $x = (x_1, x_2, \ldots, x_D)$ which belongs to the $D-1$ dimensional simplex and $x_D = 1 - \sum_{i=1}^{D-1} x_i$

$$f(x) = \sum_{i=1}^{M} \rho_i \sum_{j=1}^{n_i} \pi_{ij} \frac{1}{B(\alpha_{ij})} \prod_{k=1}^{D} x_k^{\alpha_{ijk}-1}$$

CLASS 1

⋮

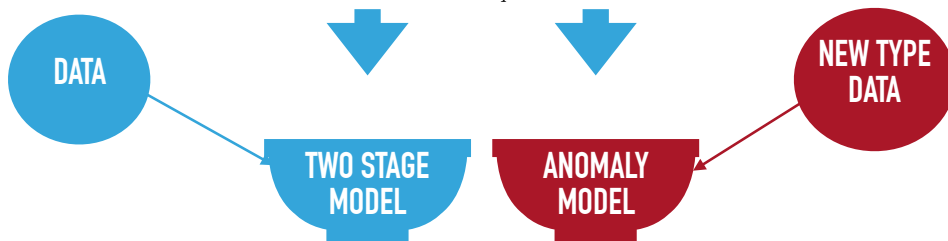**CLASS 23**

⋮

TWO STAGE MODEL
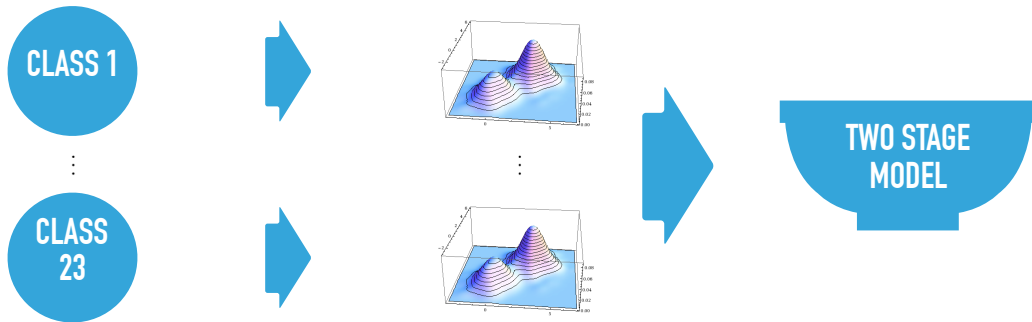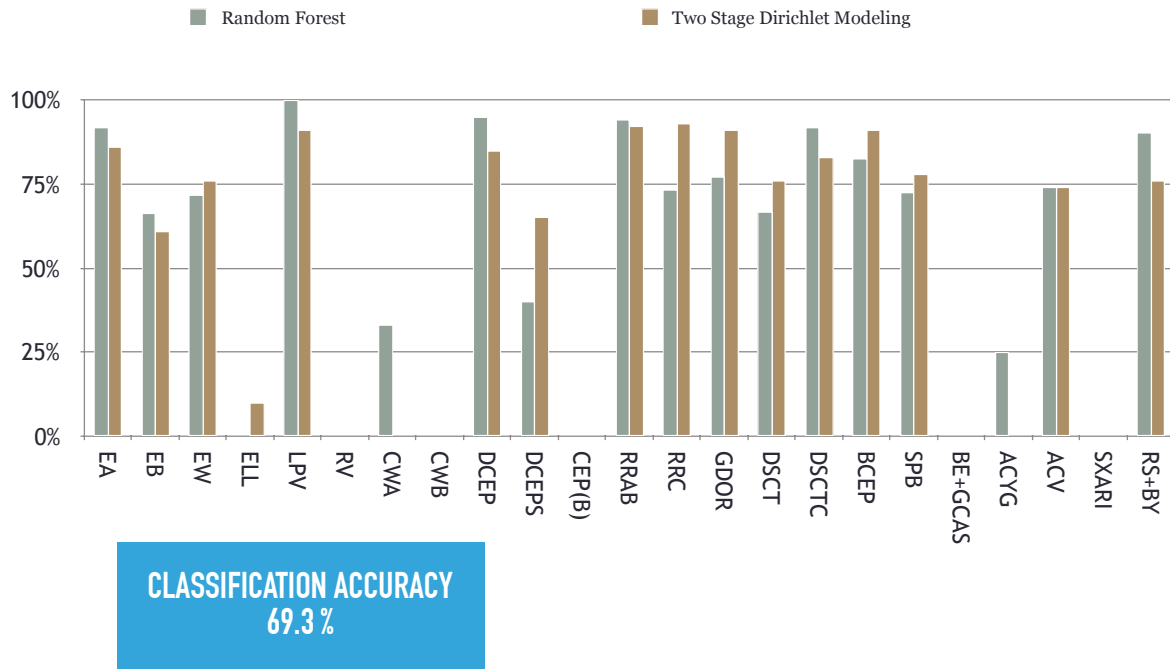
GOAL 2 : NEW CLASS DETECTION

We propose the use of a semi-supervised classification method (Vatanen et al. 2012) for detecting new classes of variable stars, called the **Fixed Background model.**

Again for the D dimensional dataset $x = (x_1, x_2, \ldots, x_D)$ which belongs to the $D-1$ simplex and $x_D = 1 - \sum_{i=1}^{D-1} x_i$,

$$f_{FB}(x) = \pi_B p_B(x) + \sum_{q=J-1}^{J+Q} \pi_q Dir(x|\alpha_q)$$

DATA

TWO STAGE MODEL

ANOMALY MODEL

NEW TYPE DATA

RESULTS : CLASSIFICATION

■ Random Forest   ■ Two Stage Dirichlet Modeling

**CLASSIFICATION ACCURACY
69.3 %**

RESULTS : NEW CLASS

■ Two Stage Classification accuracy   ■ New Class Detection Sensitivity
■ New Class Detection Specificity