



**Application Number / Title:** 22501 - Interactive Ensemble clustering for mixed data with application to mood disorders

**Applicant PI:** Assistant Professor Ellen Eischen

**Applicant Institution:** University of Oregon, Department of Mathematics, Fenton Hall, Eugene OR 97403, United States

**Lead Collaborators:**

- 1) Professor David Gotz
- 2) Professor Brian Chapman
- 3) Mathews Jacob
- 4) Dr Rachael Hageman Blair

**Collaborating Institutions and Addresses:**

- 1) University of North Carolina at Chapel Hill, Carolina Health Informatics Program, 206 Manning Hall, CB 3360, Chapel Hill NC 27599, United States
- 2) University of Utah, Radiology and Imaging Sciences,UCAIR, 729 Arapeen Drive, Salt Lake City UT 84108, United States
- 3) University of Iowa, Electrical and Computer Engineering, 4016, Seamans Center, Iowa City IA 52241, United States
- 4) University at Buffalo, Department of Biostatistics, 709 Kimball Tower, 3435 Main Street, Buffalo NY 14214, United States

**Funding body:** National Science Foundation Division of Mathematical Sciences grant

**Keywords provided by the Applicant PI to describe the research project:** depression, mood disorders, clustering

**Application Lay Summary:**

**1a:** The aim of our project is to develop an innovative approach that utilizes large population datasets to improve the classification and diagnosis of mood disorders. This project is motivated by the hypothesis that the development of novel clustering methods will enable the identification of clinically significant structures with these large population data sets. Such an approach must overcome a large number of methodological challenges introduced by the complexity of the problem and the nature of large-scale health data, including complex and unknown structure, high dimensionality, heterogeneity, and a complex mixtures of variables

**1b:** The psychiatric community has recognized the critical need for a more precise, evidence-based approach for the diagnosis and treatment of mental illnesses. Mood disorders are the leading cause of disability worldwide; about 1 in 5 British adults will experience depression at some point in their lives. The proposed program paves the way for this vision by developing new algorithms and visual tools for precision classification and diagnosis. The rigorous identification of subgroups of individuals within heterogeneous populations will facilitate accurate and targeted diagnosis, and provide opportunity for personalized evidence-based interventions.

**1c:** We are developing a novel methodology for clustering that is based on a consensus approach that accounts for uncertainty in the population and clustering method. Various components of the project will entail methodology development, data cleaning and processing, application of methods to data, and interpretation of results. Different tasks require different expertise, all of which are present in the team, but each task is motivated and will connect to the data requested from the UK Biobank. The research will be undertaken at our various institutions, and we connect in bi-weekly meetings to discuss and facilitate progress.

**1d:** We are requesting access to the full cohort data. The Scientific Team has agreed that it is acceptable to tweak the wording if necessary to rectify any words spelt incorrectly etcetera. The Applicant PI has been made aware.