Master thesis project advertisement

Title:	Machine learning approaches for diagnosis of rare genetic neurodevelopmental disorders from speech.
Type:	External Master Thesis project at Radboud UMC, Department of Clinical Genetics
Period:	Academic year 2026-2027, starting date to be discussed.
EC:	30 or 45
External supervisors RUMC:	Dr. Lex Dingemans, Prof. Bert de Vries
Internal supervisor AI:	Dr. Kiki van der Heijden

Description:
Diagnosis of rare genetic neurodevelopmental disorders such as Intellectual Disability (ID, prevalence of 2–3% of the population or ~100–200 million people worldwide) remains challenging. Despite advances in gene sequencing technologies, in up to 30 % of cases in which a genetic variant is identified, the clinical significance of the variant remains unclear, blocking diagnosis and adequate medical care (Figure 1). Clinicians therefore seek to interpret such "Variants of Unknown Significance" (VUS) by looking at a patient's phenotype such as their facial features. In recent years, machine learning approaches have been successfully utilized to recognize such facial features in an automated manner to support the assessment of VUS [1]. While these results are promising, 60% of individuals with a VUS lack recognizable facial features. Other promising phenotypic features that can contribute to the diagnosis of ID are speech and language patterns, but these remain underexplored.

The present project therefore investigates to what extent we can distinguish rare genetic syndromes based on speech recordings using machine learning. You will develop a classification pipeline combining interpretable speech features with speech embeddings derived from pre-trained speech models in a low-data regime and evaluate performance in a rigorous manner, accounting for potential confounders such as age and gender.

Who are we looking for?
We are looking for a Master's student in Artificial Intelligence (or a related program) with a solid foundation in machine learning and Python. You enjoy working with data, set up reproducible experiments, and communicate clearly about assumptions, results, and limitations. Affinity with audio, speech, or deep learning is a plus, but most importantly, you find it interesting to translate clinical questions into substantive, technical work.

What do we offer?
A thesis project focused on a clinically relevant AI case with clear societal impact. You will work in a multidisciplinary team consisting of medical experts in clinical genetics (RadboudUMC) and data-driven machine learning methods related to speech and language (Radboud University). You will have access to unique speech data of patients with rate genetic neurodevelopmental disorders and receive guidance on both the technical implementation and the clinical interpretation. If the results are convincing, the work may be further developed into a larger study and (co-)publication.

Start date and duration are negotiable depending on the choice for a 30 EC or 45 EC master thesis project. Because speech recordings may contain sensitive personal data, you will work within agreed frameworks for data access and storage.

Applications

You can apply by sending a short motivation letter (maximum half a page) and your CV to Kiki van der Heijden (kiki.vanderheijden@donders.ru.nl). Please include your availability and relevant experience (for example, projects involving speech, machine learning, or statistics).

References:

Dingemans, A.J.M., Hinne, M., Truijen, K.M.G., Goltstein, L., van Reeuwijk, J., de Leeuw, N., Schuurs-Hoeijmakers, J., Pfundt, R., Diets, I.J., den Hoed, J., et al. (2023). PhenoScore quantifies phenotypic variation for rare genetic diseases by combining facial analysis with other clinical features using a machine-learning framework. Nat. Genet. *55*, 1598–1607.