ERA Operative Headquarters
Strada dei Mercati 16/A
43126 Parma, Italy
Tel: +39 0521 989078
fellowships@era-online.org
www.era-online.org

ERA Long-Term Research Fellowship Project

CKD-MBD

Project's key info

Title of the project	**A novel artificial intelligence-powered nuclear magnetic resonance (NMR)-based blood test to predict coronary artery calcification**
Working Group involved in the project	Chronic Kidney Disease – Mineral and Bone Disorder Working Group (CKD-MBD)
Principal Investigator(s) of the project	Martin H. de Borst (The Netherlands)
Duration	12 months
Fellowship Grant	34.495,00 €
Start of the fellowship	Within 6 months after notification of the grant award to the fellow.

Receiving Institute

Name of receiving institute	University Medical Centre Groningen, The Netherlands
Supervisor's name	Martin H. de Borst (The Netherlands)
Supervisor's e-mail address	m.h.de.borst@umcg.nl

Project's detailed description

Project description
This project provides a unique opportunity to create a powerful prediction algorithm based on clinical, biochemical, genetic, nutritional data and NMR metabolomics data and to generate key insights into determinants of coronary artery calcification in individuals with and without CKD and diabetes. The primary aim of the project is to develop a novel NMR-based, AI-driven algorithm that predicts coronary artery calcification. The second aim is to pinpoint the main clinical, biochemical and nutritional determinants of CAC. This will be made possible by the explainable AI methods. Finally, as the cohort consists for 10% of patients with CKD and for 10% of patients with diabetes, with considerable overlap, this allows to also analyse these subgroups specifically. Thus, the third aim is to compare the main drivers of CAC among individuals with CKD, individuals with diabetes and individuals with both diabetes and CKD. In case of successful development of the algorithm, cost-effectiveness will also be evaluated to prepare for clinical application. Research plan Available cross-sectional data, specifically coronary calcification scores (Agatston and volume scores), anthropometric, clinical, biochemical, genetic and lifestyle data from ImaLife (up to N=12,000) will be utilized. While the current project focuses on cross-sectional analyses, we aim to secure additional funding for longitudinal analyses in a follow-up project. All cases with available data will be used to train and internally validate the machine learning model.

The primary outcome that we will predict is the Agatston score, a score based on the extent of coronary artery calcification detected by an unenhanced low-dose CT scan. Additionally, we will develop models to predict other endpoints: calcium volume score, calcium score per coronary artery, and number of calcifications per coronary artery.
First, to select the optimal number and combination of features from both the NMR spectra and the available clinical, laboratory and lifestyle data, feature selection will be done. Three different feature selection methods will be employed: FeatBoost, Boruta, and feature selection based on XGBoost feature importances. Several algorithms will be tested for performance to construct the machine learning model, such as random forests, XGBoost and bagging models. The selected features with the best-performing machine learning method will subsequently be used to perform hyperparameter tuning using 10-fold cross-validation. The best combinations of features and hyper-parameters will be used for model training on the whole training set. Internal evaluation will be done on 20% of the dataset which was not used during training. Additionally, dimensionality reduction will be performed by principal component analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) for data visualisation.
All required data including NMR data will be available at the start of the project.

The schedule for the project is as follows:

- Months 1-3: Data preparation and cleaning
- Months 4-7: Construction and evaluation of prediction models in full population and subgroups (CKD, diabetes and combined) and eventual cost-effectiveness analysis.
- Months 8-12: Writing and submission of results to national and international congresses (preferably ERA Annual Congress) and scientific (open access) journals.

Goals of the project

The project aims to:

- develop a novel NMR-based, AI-driven algorithm that predicts coronary artery calcification in the ImaLife cohort;
- identify (potentially actionable) key determinants of coronary calcification, based on the rich available data source in the Lifelines/Imalife cohort;
- perform subgroup analyses in individuals with CKD, individuals with diabetes and individuals with both diabetes and CKD, and compare the main drivers of CAC in these subgroups.

Qualifications and/or expertise required to the fellow

Qualifications of the candidate that are required for successful execution of the project are:
a background in clinical data science with specific skills in machine learning/advanced epidemiology/ modelling. Experience with cost-effectiveness analysis is desirable.