Machine Learning-Based Prediction of Cardiovascular Disease Risk

A 5-Year Forecast Using 22 Million Data Points from Nordic Countries and France

Arti Rawat[1], Julien Bezin[1], Adrian G. Zucco[2], Rachel B. Forster[5], Tibor V. Varga[2], Andrea Ganna[4], Anna-Leena Vuorinen[3,4] and Gayo Diallo[1*]

Contact: gayo.diallo@u-bordeaux.fr

Introduction

This study is one of the five use cases selected by the **European Health Data Space (EHDS)** pilot program, serving research, innovation, policy-making and regulatory purposes.

The EHDS is a regulatory framework designed to standardize health data use in the European Union by establishing a cross-border infrastructure and secure, collaborative ecosystem.

Objective: Compare cumulative incidence and estimate **cardiovascular disease (CVD) risk** using machine learning models across France, Denmark, Finland, and Norway.

Methods

Data Collection

- **22 million individuals** aged 18-85 years
- National health registries data: 2010-2018
- **4 countries**: France, Denmark, Finland, Norway

CVD Events Analyzed

- Ischemic heart disease
- Myocardial infarction
- Angina pectoris
- Ischemic stroke

Machine Learning Model

Gradient Boosted Decision Tree (GBDT) trained on all diagnoses and medication records to predict 5-year CVD risk (2014-2018)

Cohort Characteristics

France
n = 11,744,558
Female: 53.9%
Mean age: 46.95 ± 17.32

Finland
n = 4,145,551
Female: 50.9%
Mean age: 48.5 ± 17.6

Denmark
n = 4,180,434
Female: 50.6%
Mean age: 47.42 ± 17.24

Norway
n = 3,636,535
Female: 49.9%
Mean age: 46.74 ± 17.23

References

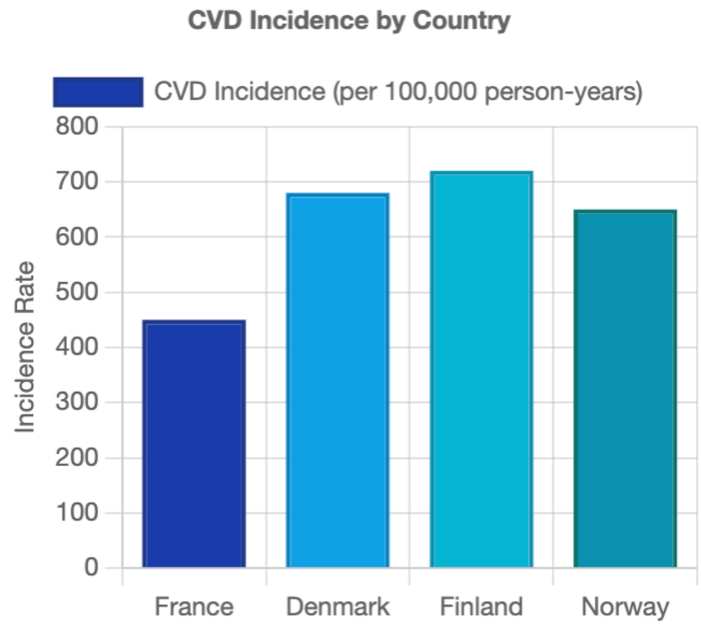
1. European Commission. Communication from the commission - a european health data space: harnessing the power of health data for people, patients and innovation. COM(2022) 196/2.
2. Ganna A, et al. The European Health Data Space can be a boost for research beyond borders. Nat Med. 2024;30(11):3053-3056.

[1]Team AHeaD BPH Inserm 1219, Univ. Bordeaux, F-33000, Bordeaux, France
[2]Copenhagen Health Complexity Center, Department of Public Health, University of Copenhagen
[3]Unit of Health Sciences, Faculty of Social Sciences, Tampere University, Tampere, Finland
[4]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland
[5]Department of Health Registry Research and Development, Norwegian Institute of Public Health, Bergen, Norway

Conference: DPH 2025

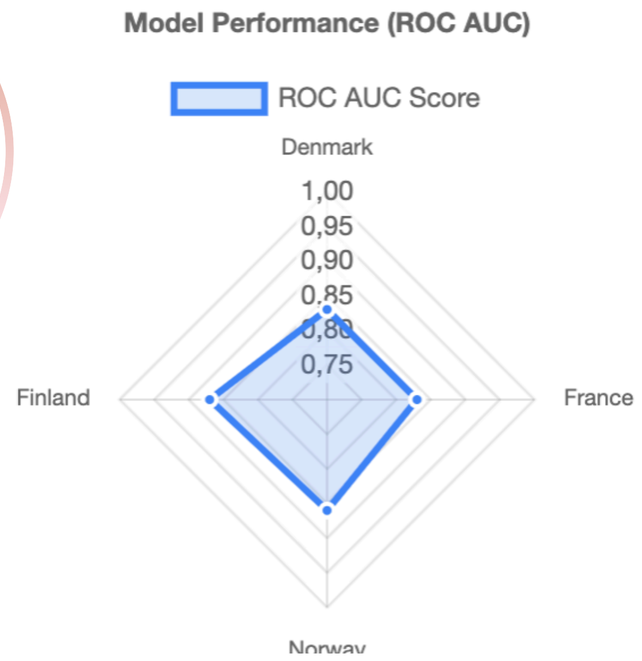
Results

CVD Incidence Comparison



France shows significantly lower CVD incidence rates compared to Nordic countries, indicating better CVD prevention, lifestyle, and health services.

GBDT Model Performance



ROC AUC Scores
Denmark: 0.83 |
France: 0.83
Norway: 0.86 |
Finland: 0.87

Brier Scores
France: 0.02 |
Finland: 0.04
Norway: 0.04 |
Denmark: 0.07

Conclusions

This study provides preliminary understanding of cross-country comparability, guiding future research and targeted investigations.

Key Findings

Comparable Model Performance: GBDT achieved similar accuracy across all countries (ROC AUC: 0.83-0.87)

Regional Differences: France showed significantly lower CVD incidence rates

Future Research Directions

- Incorporate time-varying covariates for medication usage
- Address data imbalance using sampling techniques
- Enhance cross-border health data sharing initiatives