Frederik Steensgaard Gade

Copenhagen, Denmark fstga@dtu.dk fzsg@novonordisk.com +45 5358 1612 fsgade
FSGade Google Scholar 0000-0001-9717-5242

I'm an Industrial PhD Student at Novo Nordisk R&ED and DTU. During my studies and participation in multiple research projects, I have gathered experience with several bioinformatics & AI methodologies applied to life science data.

The research projects cover multiple fields within bioinformatics, including:

- Childhood cancer genomics
- Structure-based B-cell epitope prediction
- Bayesian genotype-phenotype correlation
- Machine-learning (ML)-based polygenic risk scores for obesity, and
- Explainable disease progression prediction from electronic health record data

I am deeply passionate about leveraging AI methodologies to extract comprehensive insights from big data in life science, for the purpose of enhancing target discovery, deepening disease understanding, and ultimately improving patient outcomes.

Education

PhD — **Novo Nordisk A/S & Technical University of Denmark** — May 2024 –
AI & Digital Research & Health Bioinformatics and Personalized Medicine

- Integrating large language models and knowledge graphs for task-specific question answering in the life sciences
- Utilising patient-centric evidence from human omics datasets to help understand disease

MSc(Eng) — **Technical University of Denmark**, Bioinformatics and Systems Biology — Sep 2018 – Jul 2023

- Combined BSc(Eng) (Life Science Engineering) and MSc(Eng) (Bioinformatics and Systems Biology) on the Honours Programme
- **GPA:** 11.4/12 (*no class ranking or distinction is given by the university*)
- **Selected coursework:** Deep Learning, Protein Structure and Computational Biology, Computational Precision Medicine, Algorithms in Bioinformatics
- **Extracurricular activities:** Student council member, Student house volunteer, Young Scientists (Danish national science fair) jury and volunteer, Teaching assistant in Introductory Chemistry for Life Sciences

Experience

Novo Nordisk A/S, Junior Research Scientist — Måløv, Denmark
Strategic Development & Digital Transformation, Global Drug Discovery, R&ED — Aug 2023 – May 2024

- Matured and shaped both external and internal collaborations with *in silico* components in an effort to address current needs across therapy areas (such as obesity, diabetes, and cardiovascular disease).
- Supported research projects on an ad-hoc basis (omics analyses, data insights, etc.)

Novo Nordisk A/S, Student Intern — Måløv, Denmark
Digital Science & Innovation, R&ED — Oct 2021 - Oct 2022 & Jan 2023 - Jul 2023

- Benchmarked a variety of classification models for irregularly sampled multivariate time series on electronic health records (EHR) to predict progression of disease (e.g. diabetes to chronic kidney disease)
- Developed machine-learning (ML)-based disease risk scores based on patient genetics (ML-based polygenic risk scores)
- Stratified disease progression using using post-hoc explainability measures, such as Integrated Gradients and SHAP
- Applied clustering methods and unsupervised stratification approaches, such as functional PCA on disease trajectories
- Utilised knowledge graph embeddings in an effort to incorporate existing biomedical knowledge
- Worked extensively with the UK BioBank dataset, a large, deeply phenotyped population with multi-omics and EHR linkage

Technical University of Denmark, Student Assistant — Kgs. Lyngby, Denmark
Disease Data Intelligence, Section for Bioinformatics — May 2020 – Apr 2022

- Clinical collaboration with the STAGING project at Rigshospitalet
- Explored the role of germline and somatic variation in childhood acute lymphoblastic leukemia using clustering of mutation signatures and gene set annotations
- Called structural variants in children with acute lymphoblastic leukaemia
- Set up and ran WGS & RNA-Seq pipelines

Publications

Predicting Multimorbidity in 50,000 People Living with Obesity: A Machine Learning Model Applied to Understand Obesity Progression in Two Health Care Systems — Jan 2025

Line Egerod, Rikke Linnemann Nielsen, Zahra McVey, Joseph Katigbak, Thomas Monfeuga, ***Frederik Steensgaard Gade***, August T. H. Schreyer, Luis G. Leal, William G. Haynes, Alex Greenfield, Ella Nkhoma, Robert R. Kitchen, Michael L. Wolden, Kasper S. Matthiessen, Laurent Gautier, Abd A. Tahrani, Ramneek Gupta

10.2139/ssrn.5095147
Preprint (SSRN)

Data-driven identification of predictive risk biomarkers for subgroups of osteoarthritis using interpretable machine learning — Apr 2024

Rikke Linnemann Nielsen, Thomas Monfeuga, Robert R. Kitchen, Line Egerod, Luis G. Leal, August Thomas Hjortshøj Schreyer, ***Frederik Steensgaard Gade***, Carol Sun, Marianne Helenius, Lotte Simonsen, Marianne Willert, Abd A. Tahrani, Zahra McVey, Ramneek Gupta

10.1038/s41467-024-46663-4
Nature Communications

DiscoTope-3.0: improved B-cell epitope prediction using inverse folding latent representations — Feb 2024

Magnus Haraldson Høie, ***Frederik Steensgaard Gade***, Julie Maria Johansen, Charlotte Würtzen, Ole Winther, Morten Nielsen, Paolo Marcatili

10.3389/fimmu.2024.1322712
Frontiers in Immunology

Current projects

An overview of the projects I am currently working on during the PhD (note that due to the nature of the Industrial PhD, the project code is not shareable before publication).

Zero-shot biomedical knowledge graph construction

- Benchmarking large language models capabilities in extracting knowledge graph triplets from biomedical, scientific texts given a schema
- Tools Used: Python (`ms-swift`, `deepspeed`, etc.), R, Bash

Bariatric surgery trial multi-omics analysis

- Analysing a deeply phenotyped cohort (longitudinal proteomics, metabolomics, and more) for the purposes of understanding the effects of bariatric surgery-induced weight loss
- Tools Used: R (`limma`, `navmix`, etc.), Python (`alphapept`)

Body composition of weight loss (WL) in UK Biobank

- BMI-matched analysis of WL and non-WL UK Biobank participants and their body composition + additional life-style factors
- Tools Used: R

Theses

MSc Thesis: "Predicting and stratifying disease progression using deep learning approaches on UK Biobank longitudinal patient data" Jan 2023 - Jul 2023

- In collaboration between Novo Nordisk A/S and The Technical University of Denmark
- Grade: 12 (US: A+)

BSc Thesis: "Bayesian Inference of Amino Acid Level Geno-/Phenotype Correlation in Protein Multiple Sequence Alignments" Feb 2021 - Jun 2021

- Grade: 12 (US: A+)