SVEN DEGROEVE

PhD, Artificial Intelligence Expert

sven.degroeve@gmail.com
+32 (0)485 02 91 36
2018 Antwerpen
27-02-1977
linkedin.com/in/svendegroeve
github.com/sdgroeve
kaggle.com/svendegroeve

With over two decades of experience in Artificial Intelligence, I have successfully navigated the intersection of research and industry, leveraging AI to address complex, real-world challenges. My expertise spans statistics, data analysis, machine learning, and deep learning, with a proven ability to translate AI innovations into impactful solutions in areas such as natural language processing, biotechnology, and biomedical predictive modeling. While my accomplishments in the biotechnology domain underscore my technical proficiency, I am eager to broaden my horizons and apply my AI expertise to new industries and challenges. I quickly absorb and implement new concepts, and I excel at articulating complex ideas in a clear, accessible manner. As a team player, I thrive in dynamic environments and effectively bridge the gap between technical depth and practical application. With the rapid advances in AI, I'm excited to contribute my problem-solving skills, innovative thinking, and strong communication abilities in an industry-driven role that spans multiple sectors, driving both product development and business growth.

PROFESSIONAL CAREER

Associate Professor in Machine Learning for Biomedical Data, Senior Lecturer
Oct 2019 – Present | Faculty of Medicine and Health Sciences, Ghent University

My significant contributions to the integration of machine learning and AI in both education and research within the Faculty of Medicine and Health Sciences resulted in a 10% Associate Professorship to further advance my educational initiatives. These are highly regarded by both the faculty and master's students. Additionally, I serve as (co-)promoter for pre- and postdoctoral researchers. I teach students all about the current state-of-the-art in machine and deep Learning as well as generative AI. Students learn to use Scikit-learn, PyTorch and Transformers. The course link is provided below.

Since 2021, I have served as a member of the Educational Committee for Biomedical Sciences and the IT Committee, supporting the faculty in its strategic preparation for the AI-driven future. I'm a member of the Health Intelligence Network Ghent, a multidisciplinary initiative that brings together expertise in AI and healthcare.

Senior Staff Scientist, AI Expert (AI biotechnology, high-throughput proteomics)
Mar 2017 – Present | CompOmics, VIB-UGent Center for Medical Biotechnology, Zwijnaarde

I lead and promote machine and deep learning research and development within the Computational Omics and Systems Biology Group (CompOmics) of the Flemish Institute for Biotechnology (VIB). My research findings are currently being integrated into Bruker Corporation's commercial software to enhance the analysis of mass spectrometry data. In parallel I lead the development of a SaaS called ionbot™ funded by an Industrial Research Fund (IOF) grant of Ghent University. This tool implements

optimized classical machine models that accurately predict the behaviour of molecules in proteomics experiments.

As a Staff Scientist I also apply my expertise to advance AI-based data analysis for other groups within bot the VIB center and the Ghent University Hospital. This includes the application of Large Language Models for protein analysis (e.g. ESM2), predictive modelling for plant gene regulation understanding and bacterial resistance prediction.

Machine Learning Expert (AI biotechnology, computational biology)
Jul 2009 – Feb 2017 | CompOmics, Medical Biotechnology Center, VIB, Ghent University

As an early member of a small bioinformatics group, I was responsible for data analysis in a wide range of proteomics experiments within the center. Over time, this group expanded into CompOmics, now a thriving research team of over twenty members with a strong reputation for developing cutting-edge research tools and producing high-impact publications. I played a pivotal role in securing high-budget grant funding and attracting top talent to the group by integrating machine learning and AI into the Master's program in Biomedical Sciences.

Senior Knowledge Systems Engineer (clinical trial optimization AI software)
Oct 2008 – Mar 2009 | MDCPartners, Antwerp

MDCPartners was a small start-up that provided business intelligence solutions for clinical trial optimization, medical expert identification and competitive intelligence. I investigated the application of machine learning methods to further improve on the optimization software in development.

Data Analyst, Senior Scientist (biomarker discovery, statistics)
Oct 2005 – Aug 2008 | Pronota, Technologiepark Zwijnaarde (now part of Biocartis, Mechelen)

Pronota was a spin-off company of the Flemish Institute for Biotechnology (VIB) that aimed to build a portfolio of diagnostic products that work by analyzing protein biomarkers in body fluid using high-throughput proteomics experiments. I was responsible for the experimental design and analysis of the proteomics studies for both discovery platform development and biomarker discovery (Python, R, Spotfire). My task was to evaluate and communicate key issues, derived from data analysis, affecting biomarker discovery and data quality.

Post-doctoral Fellow (audio signal AI)
Jan 2005- Sep 2005 | KERMIT, Ghent University

As part of an FWO funded academic research project I applied my machine learning expertise to improve the annotation of drums and snares in raw audio signals (very early WAV to MIDI conversion). My successful implementation was published and presented at the International Conference on Music Information Retrieval (ISMIR).

PhD (genomics AI revolution)
Jan 2001 –Dec 2004 | Ghent University, Flemish Institute for Biotechnology (VIB)

As part of a large CNRS International Research Network program, I implemented support vector machine based learning algorithms for the automatic annotation of newly sequenced genomes (C).

My work resulted in the first accurate gene annotation of several genomes with the results of this work published in top journals such as Science and PNAS.

Machine Learning Researcher (HIV virus mutations)
Jan 2001 – Dec 2001 | Tibotec Virco, Mechelen

As part of a newly initiated data analysis department, I introduced machine learning methods for the analysis of complex virus mutation patterns of protease resistance in HIV data. My successful proof-of-concept resulted in significant funding for the company.

Machine Learning Research Assistant (chatbots, AI dialogue systems)
Sep 1999 – Dec 2000 | Centre for Evolutionary Language Engineering (CELE), L&H, Ieper

In these early days of AI for Natural Language Processing, our blue-sky research group investigated the application of state-of-the-art machine learning algorithms for solving Language Modeling tasks to further improve the AI dialogue systems commercialized by the company Lernout&Hauspie (L&H).

EDUCATION

Doctorate (Ph.D.) in Sciences: Computer Science
Ghent University | 2001 – 2004
Title of thesis: "Design and evaluation of a linear classification strategy for gene structural element recognition"

Licentiate (Master in Sciences) in Computer Science
Ghent University | 1997 - 1999
Title of thesis: "Classification of skin lesions with neural networks"

LANGUAGES

Dutch (native proficiency), **English** (advanced C1 level), **French** (intermediate proficiency)

SKILLS

Programming: Python, Pandas, C, Cython, R, Visual Studio Code, SQL, PostgresQL
Machine Learning: Scikit-learn, Auto-sklearn, Scikit-multilearn, Hyperopt, Kaggle
Deep Learning: PyTorch (Lightning), Tensorflow, Keras, HuggingFace API, Weights&Biases, MLFlow, AutoKeras, Google AutoML Vision
Generative AI: LangGraph
Data exploration: Seaborn, Plotly, Tableau Software
Cloud: Google Compute Engine (Vertex AI), Microsoft Azure, Kubernetes, GitHub
Other: Docker, Flask, Streamlit

TEACHING

Machine Learning methods for Biomedical Data (D012554)

Ghent University | yearly since 2016 | 55 hours
GitHub: https://github.com/sdgroeve/Machine_Learning_course_UGent_D012554_2025

Large scale analysis of Biomedical Data (D013628)

Ghent University | yearly since 2015 | 10 hours

Machine Learning & Deep Learning Workshop

Flemish Institute for Biotechnology (VIB) | twice a year since 2018 | 32 hours
Link: https://training.vib.be/all-trainings/machine-learning-deep-learning-workshop

SCIENTIFIC PUBLICATIONS

ALL PUBLICATIONS

UGent Biblio: https://biblio.ugent.be/publication?text=sven+degroeve
Google Scholar: https://scholar.google.nl/citations?user=pzwt3NYAAAAJ&hl=en

MOST RELEVANT PUBLICATIONS

M. Palmblad, S. Boecker, **S. Degroeve**, O. Kohlbacher, L. Kall, W. S. Noble, M. Wilhelm, Interpretation of the DOME Recommendations for Machine Learning in Proteomics and Metabolomics. J. Proteome Res. 21, 1204–1207 (2022).

R. Bouwmeester, R. Gabriels, N. Hulstaert, L. Martens, **S. Degroeve**, DeepLC can predict retention times for peptides that carry as-yet unseen modifications. Nat. Methods. 18, 1363+ (2021).
GitHub: https://github.com/compomics/DeepLC

R. Bouwmeester, R. Gabriels, T. Van Den Bossche, L. Martens, **S. Degroeve**, The Age of Data-Driven Proteomics: How Machine Learning Enables Novel Workflows. Proteomics. 20 (2020).

R. Gabriels, L. Martens, **S. Degroeve**, Updated MS2PIP web server delivers fast and accurate MS2 peak intensity prediction for multiple fragmentation methods, instruments and labelling techniques. Nucleic Acids Res. 47, W295–W299 (2019).
GitHub: https://github.com/compomics/ms2pip

A. S. C. Silva, R. Bouwmeester, L. Martens, **S. Degroeve**, Accurate peptide fragmentation predictions allow data driven approaches to replace and improve upon proteomics search engine scoring functions. Bioinformatics. 35, 5243–5248 (2019).
GitHub: https://github.com/compomics/ms2rescore

Zavrel, J., **S. Degroeve**, A. Kool, W. Daelemans, K. Jokinen, Diverse classifiers for NLP disambiguation tasks. Comparisons, Optimization, Combination, and Evolution. In: Jokinen et al. (eds.), TWLT 18. Learning to Behave. CEvoLE 2, Ieper, Belgium, p. 201-221, 2000.

PATENTS

Kas, K., Vanpoucke, G., **Degroeve, S.**, & Huijben, K. (2010). Biomarkers and Methods for Diagnosing, Predicting and/or Prognosing Sepsis and Uses Thereof. *U.S. Patent Application No. 20100292131*.

CERTIFICATES

Neural Networks and Deep Learning (Coursera), issued Nov 2019
Structuring Machine Learning Projects (Coursera), issued Jan 2020
Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization
(Coursera), issued Jan 2020
CEFR level C1 on the Interuniversity Test of Academic English, issued Sep 2019