ADVANCED MACHINE LEARNING IN HEALTHCARE

& LARGE LANGUAGE MODELS IN HEALTHCARE

Short Course

Date: 13 - 24 October 2025
Location: Wits University, Johannesburg
Learning type: Face-to-face

SUB-SAHARAN AFRICA CONSORTIUM FOR ADVANCED BIOSTATISTICS TRAINING **(SSACAB)**

IN COLLABORATION WITH UMC UTRECHT & WITS SCHOOL OF PUBLIC HEALTH

SSACAB
SUB-SAHARAN AFRICA CONSORTIUM FOR ADVANCED BIOSTATISTICS TRAINING

UMC Utrecht

WITS UNIVERSITY

SHORT COURSE IN ADVANCED MACHINE LEARNING IN HEALTHCARE & LARGE LANGUAGE MODELS IN HEALTHCARE

SSACAB in collaboration with the School of Public Health, Wits University and the Department of Data Science and Biostatistics, University Medical Center Utrecht will be offering two short courses in Advanced Machine Learning in Healthcare & Large Language Models in Healthcare from **13 – 24 October 2025**

WEEK 1

ADVANCED MACHINE LEARNING IN HEALTHCARE

Advanced machine learning techniques are playing an essential role in modern healthcare by aiding in the analysis of complex, high-dimensional datasets to improve diagnosis, predict patient outcomes, and personalize treatment strategies. This course is designed to equip participants with theoretical knowledge and practical skills in applying machine learning to healthcare data.

COURSE OBJECTIVES

By the end of the course, participants should:

- Be acquainted with key principles underlying main methods of machine learning, both supervised and unsupervised, such as Nearest neighbours, Bayes classifiers and discriminant analyses, Decision trees, boosting and random forest; Regularization methods (Ridge, Lasso, elastic net); Support Vector Machines; Neural networks; Principal Component Analysis and UMAP; and Two-Way Orthogonal Partial Least Squares (O2PLS).
- Be familiar with evaluating classifiers, such as ROC-curves, calibration and Bias-Variance trade-off, and supervised or unsupervised learning.
- Know the key difference between hard classification and probabilistic prediction; and is familiar with evaluation methods for probabilistic prediction.
- Has sufficient knowledge of the challenges associated with high-dimensional data.
- Is able to apply all these methods to real data and to present and interpret the results of a machine learning application to real data.

COURSE CONTENT

The course consists of five days with the following content:

DAY 1: Introduction and theoretical foundations of machine learning in healthcare

- General introduction
- Supervised learning: classification and regression
- Classification and regression in High Dimensional Data
- In-depth view: Support Vector Machines and Neural Networks

DAY 2: Advanced supervised learning techniques

- Hyperparameter tuning and model validation
- Penalized regression
- Classification and regression trees
- Bagging, random forest, boosting

DAY 3: Essentials of prediction in the clinical context

- Probabilistic prediction vs hard classification
- Evaluating predictions: discrimination and calibration assessment
- Clinical usefulness: Decision curve analysis
- Machine learning in survival analysis

DAY 4: Unsupervised learning and data integration

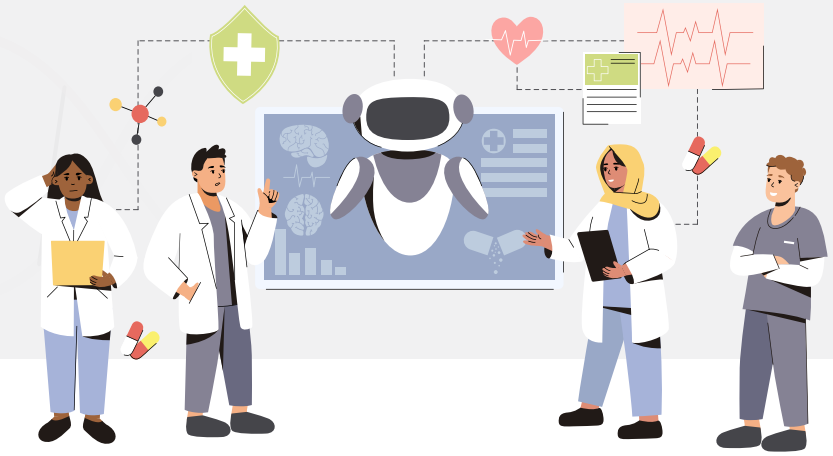
- Theoretical principles of unsupervised learning
- Machine learning for manifold approximations and clustering
- Integration of multi-modal data: Two-Way Orthogonal Partial Least Squares (O2PLS)

DAY 5: Case study assignment with presentations

SSACAB SUB-SAHARAN AFRICA CONSORTIUM FOR ADVANCED BIOSTATISTICS TRAINING

UMC Utrecht

WITS UNIVERSITY



WEEK 2

LARGE LANGUAGE MODELS IN HEALTHCARE

AI in healthcare is rapidly evolving beyond traditional machine learning methods, driven by the rise of large, general-purpose AI. These models, also called foundation models, can process diverse data types such as text, images, and structured data. This enables powerful applications in healthcare, for example in the areas of diagnostics, decision support, and clinical documentation.

COURSE OBJECTIVES

By the end of the course, participants should:

- Understand the concept and architecture of foundation models and their relevance to healthcare.
- Know the key components of large language models (LLM), vision-language models, and multimodal models.
- Are familiar with principles of transfer learning and instruction fine-tuning.
- Are familiar with the challenges of validating and implementing foundation models.
- Are familiar with various methods of evaluating foundation model performance (e.g. lexical and semantic similarity, factuality, etc.)
- Are aware of the opportunities and limitations of applying generative models to clinical practice.
- Can implement and fine-tune pretrained models for specific healthcare tasks using Python (and Google Colab, Hugging Face and Pytorch).
- Are familiar with ethical, regulatory, and interpretability challenges.

COURSE CONTENT

The course consists of five days with the following content:

DAY 1: Introduction to Foundation Models in Healthcare

- General introduction
- What are foundation models?
- Introduction to capabilities and risks medicine
- Current applications
- Examples of existing models
- General training pipeline

DAY 2: Large Language Models architecture

- Quick recap: training and principles of foundation models
- Transformer architecture essentials
- Prompting, finetuning, instruction tuning
- Hands-on: train a toy LLM on a small dataset

DAY 3: Multimodal Models

- Introduction to multimodal models
- Combining text, images, and structured data
- Vision-language alignment in radiology and pathology
- Case examples of multimodal models
- Hands-on: run multimodal models in Python or via web demos

DAY 4: Evaluation and Deployment

- Clinical tasks and associated risks
- Model explainability and interpretability
- Evaluation and failure cases of foundation models
- Evaluation methods for generative AI
- Deployment challenges: data privacy, bias, regulatory compliance

DAY 5: Group Case Study and Presentations

- Group work on use case: implement or evaluate a foundation model
- Prepare findings and practical/theoretical insights
- Choose model, task, dataset, and validation strategy
- Present to peers and receive feedback
- Course wrap-up and reflection

SHORT COURSE IN ADVANCED MACHINE LEARNING IN HEALTHCARE & LARGE LANGUAGE MODELS IN HEALTHCARE

SSACAB in collaboration with the School of Public Health, Wits University and the Department of Data Science and Biostatistics, University Medical Center Utrecht will be offering two short courses in Advanced Machine Learning in Healthcare & Large Language Models in Healthcare from **13 – 24 October 2025**

PREREQUISITES

- Some programming experience in any language, preferably in Python or R (or willingness to learn quickly).

Familiarity with basic statistical concepts (e.g. mean, variance) and preferably knowledge of the linear model/regression.

SOFTWARES

- In the first week, R will be used during the lectures to illustrate the theory.

In the second week, students will use Python during lab sessions to run and train and run foundation models.

ASSESSMENT

Daily assessment

Each afternoon there will be an interactive Q&A and a quiz. Participants may ask questions and discuss topics. Instructors will ask questions and bring in discussion points where suitable. Actively participating is mandatory.

Presentation of case study

On the last day you will work on a use case for a foundation model in healthcare in small groups (appr. 3 students), to be presented in the afternoon. The students will either finetune a new model, or evaluate an existing foundation model, for clinical purposes. The model and evaluation will be presented by the students to their peers. The case presentation will be graded according to predefined evaluation criteria that will be shared beforehand.

EVALUATION

In order to continue improving this course, we need your suggestions! At the end of the course you will be asked to fill in a course evaluation. We kindly request that you to take the time to fill in and return the evaluation form. Suggestions for improvement during the course are also welcome.

FACILITATOR'S BIOGRAPHIES

Dr. Said el Bouhaddani is Assistant Professor of Biostatistics & Machine Learning at the University Medical Center Utrecht. His research focuses on developing statistical and AI methods for complex, multimodal healthcare data, with a special interest in latent variable modeling and digital twins. His expertise is in high-dimensional multi-omics analyses involving diverse molecular data, and collaborates closely with clinical researchers on biostatistical methods in clinical trials and observational studies. Said also coordinates a new Health Data Science track at Utrecht University and teaches biostatistics and machine learning to MSc, PhD, and professional audiences.

Dr. Ruurd Kuiper is Assistant Professor at the Department of Data Science and Biostatistics at the Julius Center, University Medical Center Utrecht. His research focuses on developing and evaluating generative (language) models for clinical applications. He is particularly interested in enhancing the reasoning capabilities of large language models through alternative network architectures, and in improving the trustworthiness of foundation models through rigorous validation protocols. He also teaches, supervises, and gives presentations in the fields of machine learning, natural language processing, and AI for healthcare.

SSACAB WEBSITE

REGISTER NOW: https://forms.office.com/r/TpMCYPPMQW

For course queries email: Sbouhad2@umcutrecht.nl **or** K.Klipstein-Grobusch@umcutrecht.nl

Application Fee: **R500 ZAR**

Application deadline: 30 September 2025

SSACAB SUB-SAHARAN AFRICA CONSORTIUM FOR ADVANCED BIOSTATISTICS TRAINING

UMC Utrecht

WITS UNIVERSITY

SSACAB
SUB-SAHARAN AFRICA CONSORTIUM FOR ADVANCED BIOSTATISTICS TRAINING

UMC Utrecht

WITS UNIVERSITY