Courses in English
Course Description

Department	04 Electrical Engineering and Information Technology
Course title	**Edge Machine Learning**
Hours per week (SWS)	3
Number of ECTS credits	5
Course objective	This course is designed for Bachelor's students with a foundational understanding of Python and a strong interest in Machine Learning. By the end of the course, students will be able to explain the principles of embedded machine learning and identify the challenges and opportunities of deploying models on edge devices. They will understand key concepts in model architecture, optimization, and compression, and be able to apply techniques such as transfer learning and quantization to real-world problems. Students will also gain insights into hardware-specific considerations and the role of ML Ops in sustainable and scalable ML deployment. In hands-on labs, students will work with Raspberry Pi 5 and AI accelerators to implement and optimize ML models using Python. The exercises introduce all tools and steps required for implementing the final project, which is the major part of the "Modularbeit" that accounts for 50% of the final grade. An oral exam (50%) ensures that all learning goals of the course have been achieved. Course content: • Fundamentals of machine learning, with a focus on neural networks • Overview of standard architectures and their implementations for typical applications such as image classification, object detection, time-series analysis, and text analysis • Training and inference workflows in operational environments • Use of established frameworks (e.g. PyTorch) for model training and inference • Use and reuse of existing network modules • Retraining, fine-tuning, and transfer learning based on pre-built network architectures • Standard formats for transferring neural networks (e.g. ONNX) • Optimization of deep neural network architectures for embedded systems, including hyperparameter optimization, data type and precision reduction, optimization strategies, and embedded hardware architectures for AI applications • Model optimization tools • Deployment to the target platform
Prerequisites	Mathematics 3, Computer Engineering 3, and preferably a prior course in Machine Learning
Recommended reading	• K. C. Tung: Tensorflow 2 Pocket Reference: Building and Deploying Machine Learning Models, O'Reilly Media, Inc, USA, 2021 • A. R. Jha: Mastering PyTorch, Packt Publishing, USA, 2020 • Google: Tensorflow Lite, www.tensorflow.org/lite • Google: TFX: www.tensorflow.org/tfx • The ONNX community project: onnx.ai • Nvidia: TensorRT Developer Guide, https://docs.nvidia.com/deeplearning/tensorrt/developer-guide/index.html • Goodfellow I., Yoshua B., and Aaron C. (2016): Deep learning, MIT Press
Teaching methods	Seminar-based teaching with practice/practical training
Assessment methods	Modularbeit (final project + documentation, 50%) and oral exam (50%)
Language of instruction	English
Name of lecturer	Prof. Dr. Clemens Hage
Email	clemens.hage@hm.edu
Link	https://ee.hm.edu/kontakte_de/phonebook_detailseite_103620.de.html
Course content	See handbook for the master program "Elektrotechnik und Informationstechnik", https://ee.hm.edu/studierende/modulhandbuecher.de.html

Remarks	If you are interested in this course, please contact the lecturer to clarify your current knowledge before registering.