NVIDIA DGX systémy pro pokročilé aplikace strojového učení

Prof. Ing. Pavel Václavek, Ph.D.

23.10.2024

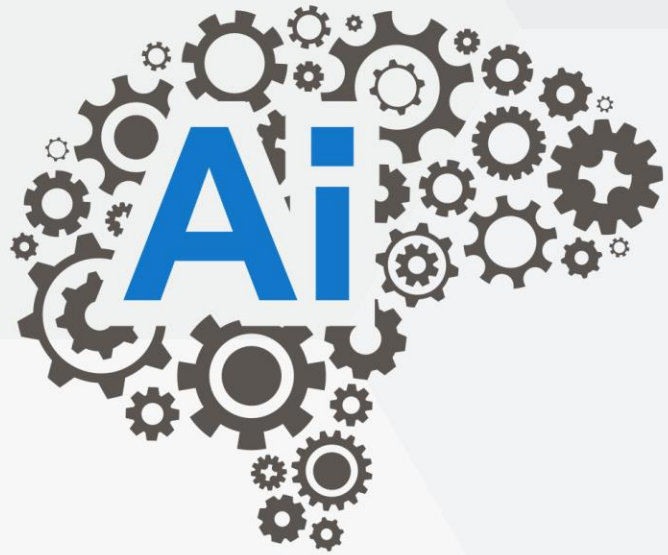
CEITEC – Central European Institute of Technology

- CEITEC založen 2011 jako evropské centrum excelence
- CEITEC VUT je vysokoškolským ústavem VUT v Brně
- Multidisciplinární institut – materiálové vědy, nano-technologie, engineering, přírodní vědy

Kybernetika a robotika

Základní a aplikovaný výzkum

- Řídicí a automatizační systémy
- Robotika – mobilní roboty
- Senzory a měřicí systémy
- Průmyslové komunikace
- Instrumentace, PLC, průmyslové řídicí systémy



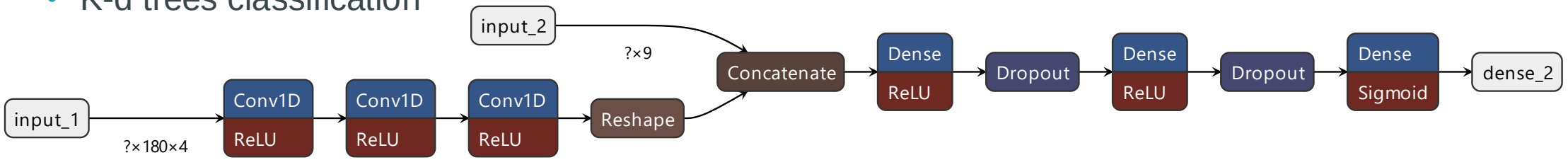
Fail-operational application

- Industry
  - Fully automatic production lines
- Automotive
  - Autonomous vehicles
- Aerospace
  - High reliable electrical propulsion
- Traction systems
  - Public transport systems

CEITEC

Fault detection

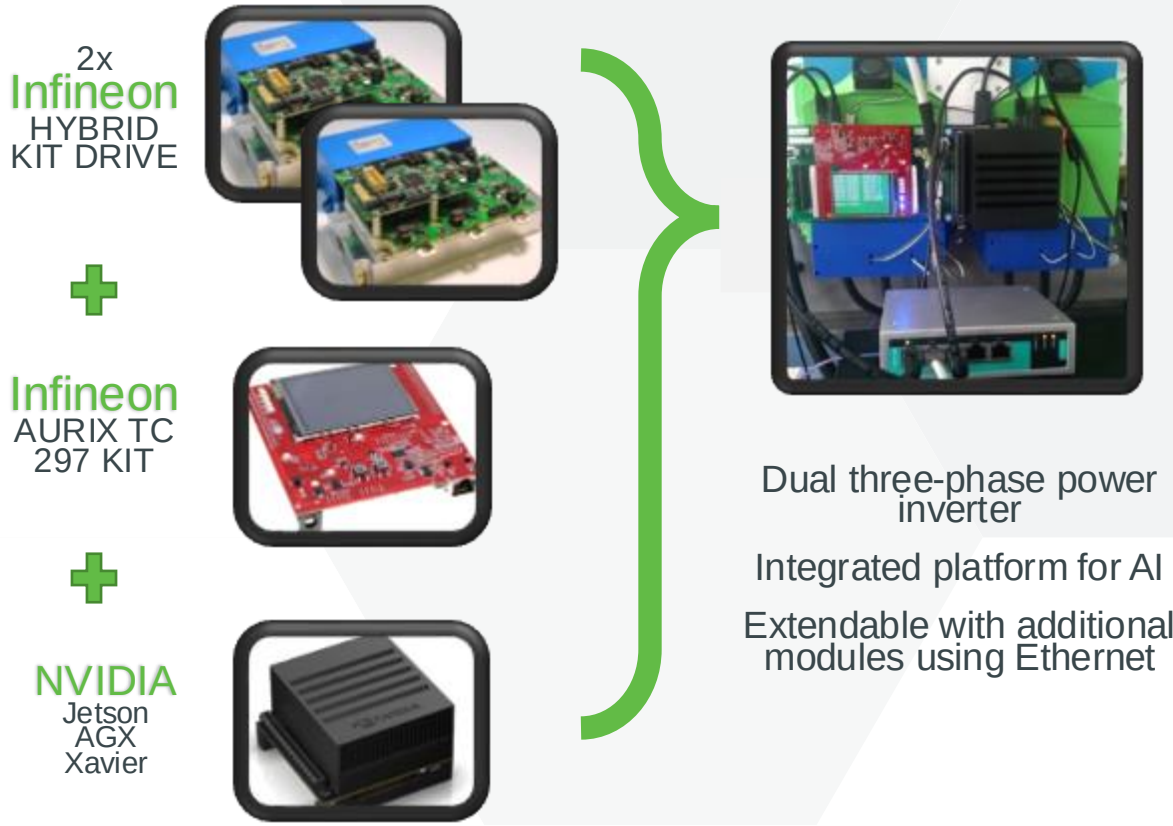
- Various faults require various detection time
- Undetected faults without appropriate action lead to other faults – fault propagation
  - Interturn short-circuit causes the temperature increase → temperature damages other coil turns of the winding
- Motor fault can be detected using various methods
  - Detection methods based on motor model
  - Parameter and state observers prepared for specific fault
  - Statistical methods
  - Machine learning methods
    - Artificial Neural Networks classification
    - K-d trees classification



CEITEC

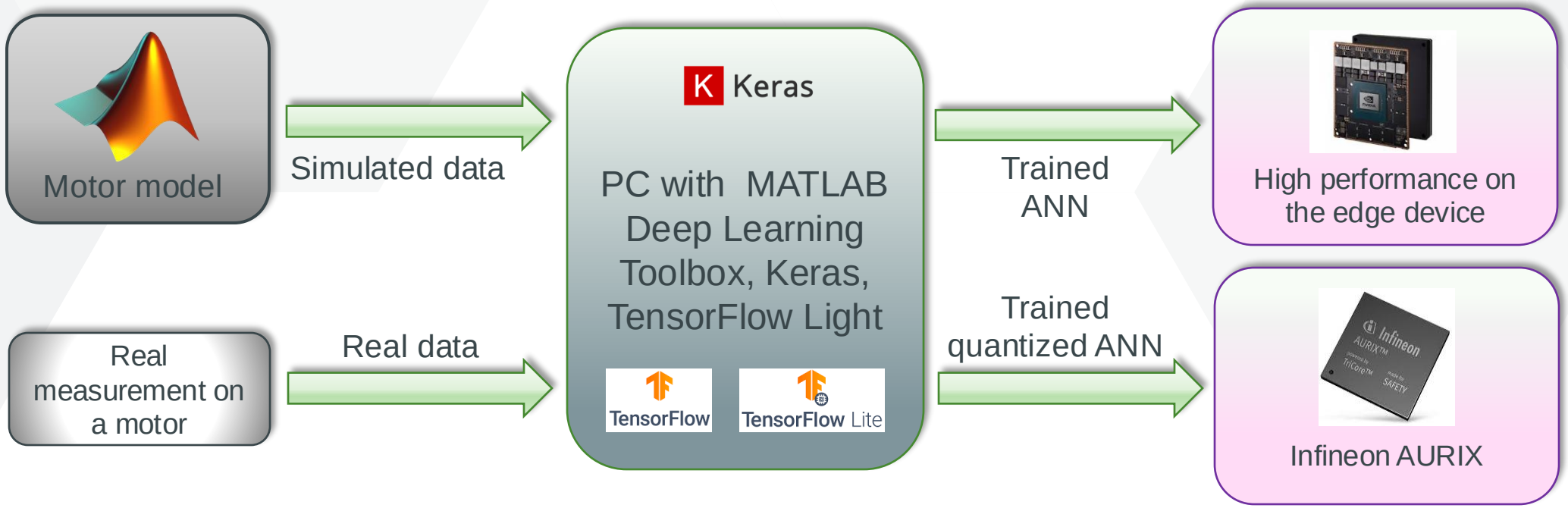
ANN experimental platform

- The power inverter consists of Hybrid kit drive, AURIX TC297 Application kit and Jetson AGX Xavier.
- Jetson AGX Xavier with 512-core Volta GPU with Tensor Cores and 8-core ARM v8.2 64-bit CPU accelerates AI computation
- AURIX and Xavier are interconnected using Ethernet
- Inverter is connected to customized configurable PMSM
- PMSM can be loaded using dynamometer
- The whole system is controlled from MATLAB



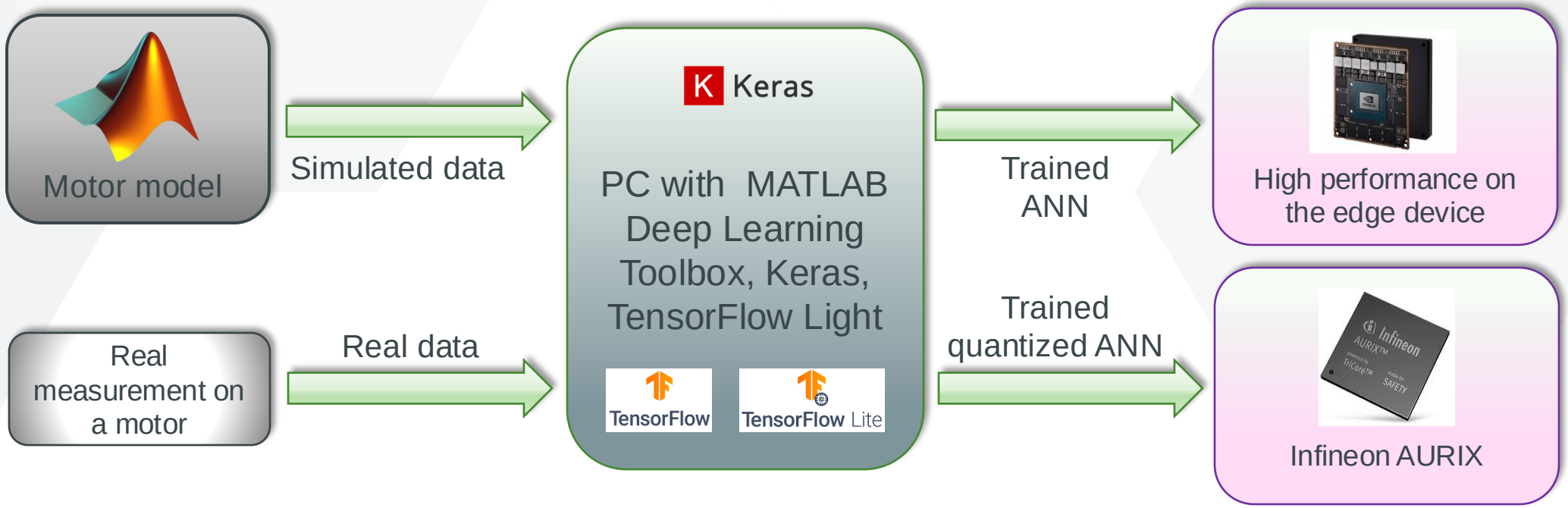
ANN preparation

- ANN is prepared in TensorFlow using combination of measured and simulated data
  - Simulated data are required for faults which can not be emulated and measured
  - Using only simulated data for training is also possible



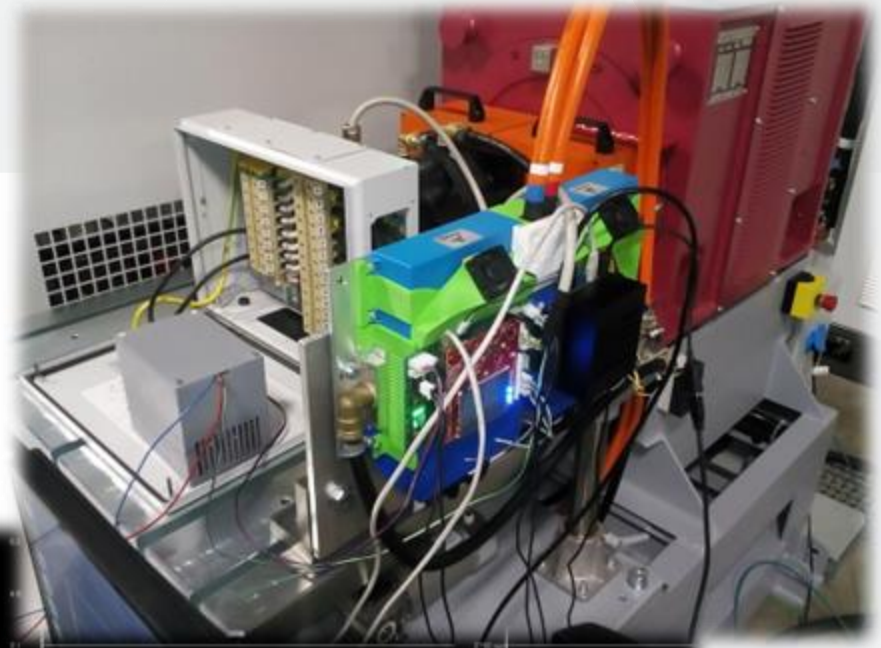
ANN preparation

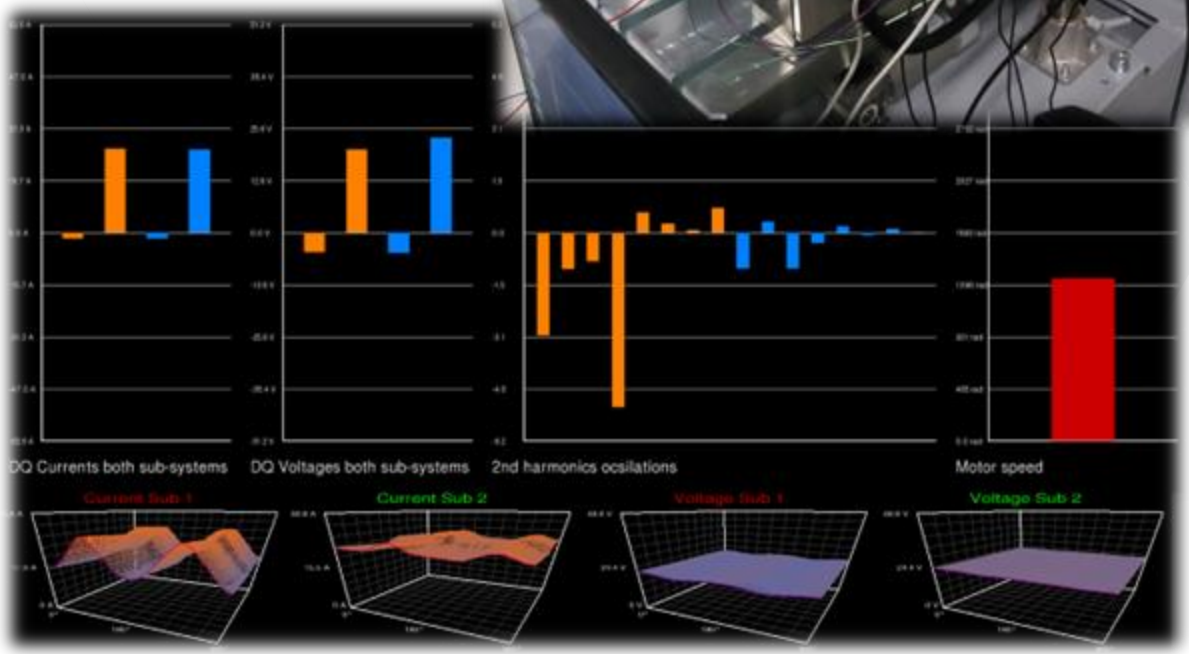
- Prepared ANN is integrated into target platform
  - Powerful AI platform can operate in **float32** or **float16**
  - Microcontrollers typically require ANN quantization into **int** for best performance

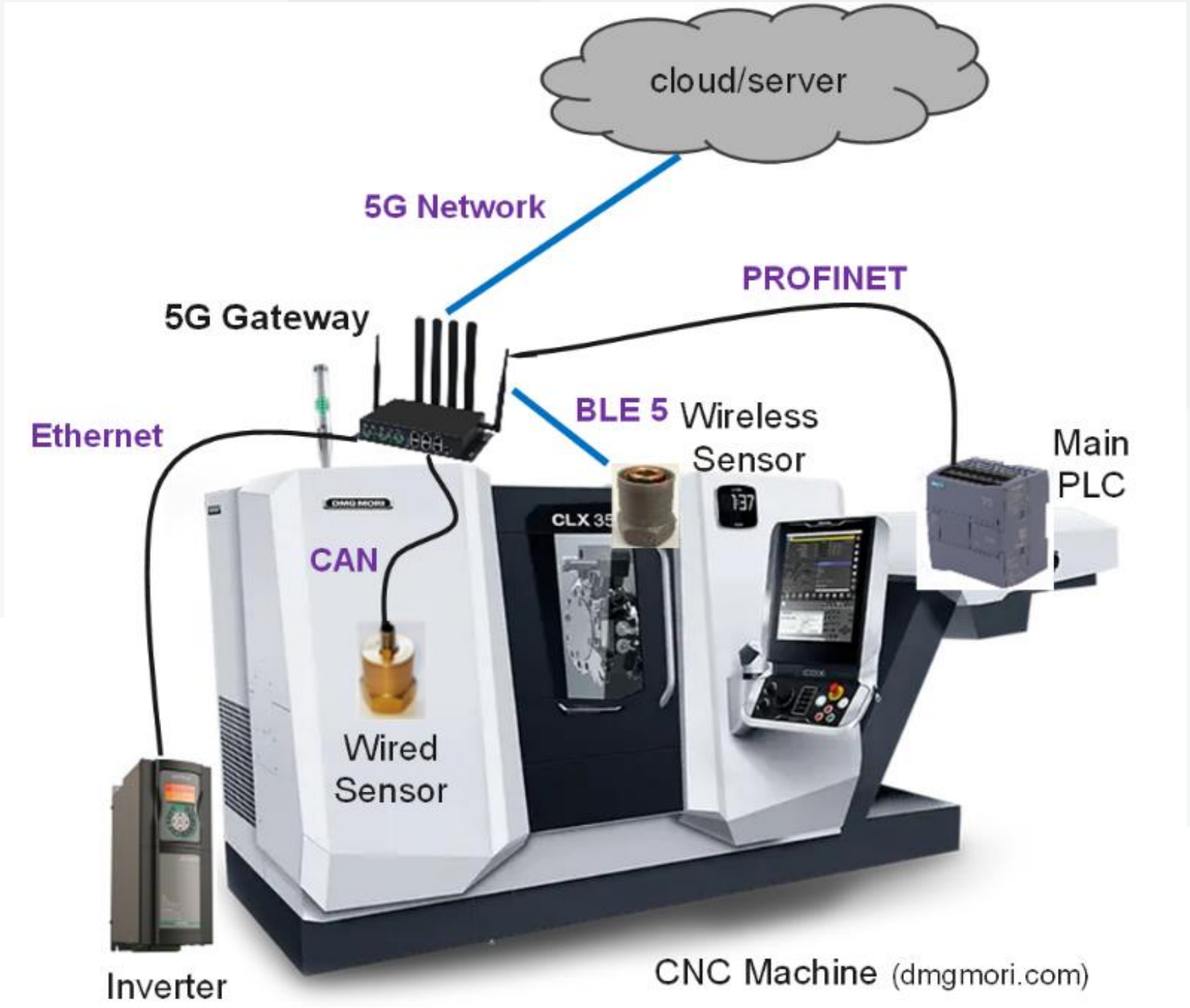


Achieved classification results

- Prepared ANN can successfully detect fault at steady state
- Classification using on the edge implementation of ANN in less than 5 ms (including data preprocessing and communication delays)
  - ANN inference time around 1ms
- Combination of convolutional and dense layers
  - One input for oversampled motor voltages and currents
  - Second input for motor speed, setpoints, and extracted features



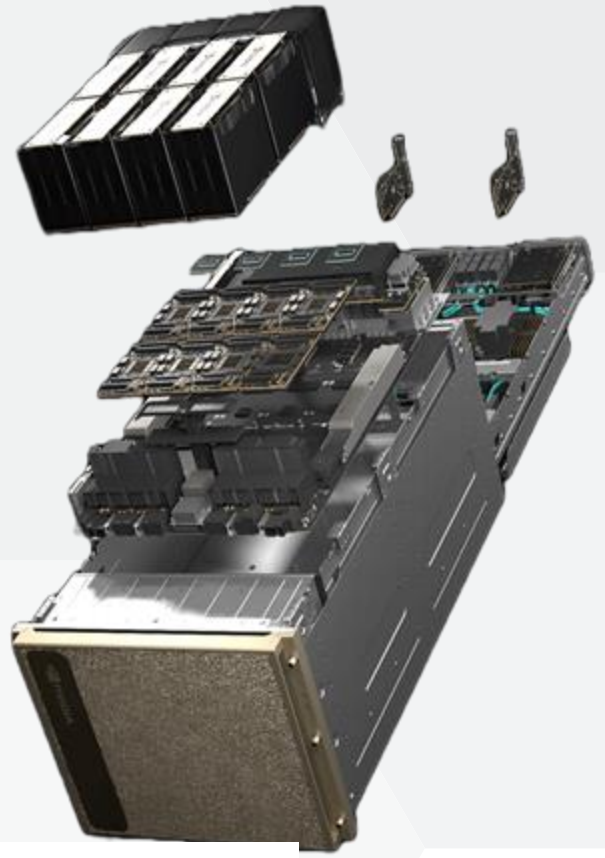


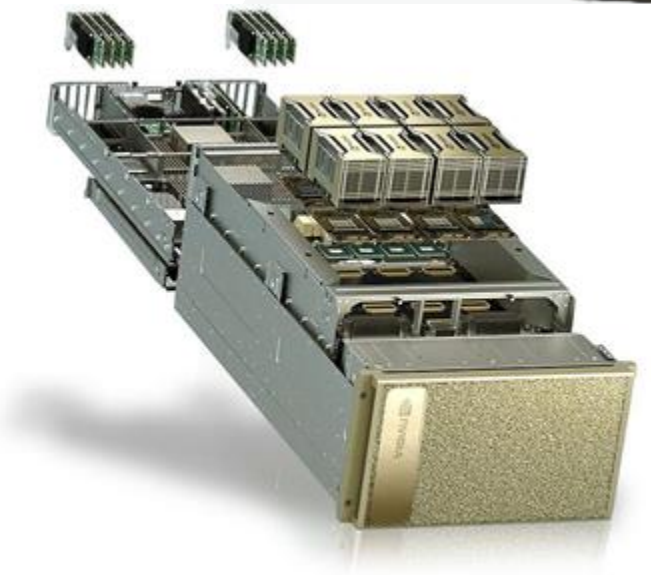

cloud/server
5G Network
PROFINET
5G Gateway
BLE 5
Wireless
Sensor
Ethernet
Main
PLC
CAN
Wired
Sensor
Inverter
CNC Machine (dmgmori.com)

CEITEC

Cluster NVIDIA DGX H100 + DGX A100

PARAMETR	NVIDIA DGX H100 640 GB	NVIDIA DGX A100 640 GB
GPUs	8× NVIDIA H100 SXM5 80 GB	8× NVIDIA A100 SXM4 80 GB
GPU memory	640 GB total	640 GB total
CPU	2x Intel Xeon Platinum 8480C CPU, (112 jader) 2.00 GHz	2× AMD Epyc 7742 (128 jader, 2.25GHz)
Výkon (tensor operace)	32 PetaFLOPS (FP8)	5 PetaFLOPS (FP16)
# CUDA jader	135 168	55 296
# Tensor jader	4 224	3 456
Multi-instantce GPU	56 instancí	56 instancí
RAM	2 TB	2 TB
HDD	OS: 2× 1.92 TB NVMe data: 30 TB (8× 3.84 TB) NVMe	OS: 2× 1.92 TB NVMe data: 30 TB (8× 3.84 TB) NVMe
Network	8x ConnectX-7 400Gb/s InfiniBand 4x ConnectX-7 200Gb/s Ethernet	8x ConnectX-7 200Gb/s InfiniBand 4x ConnectX-7 200Gb/s Ethernet
Max. spotřeba	10,2 kW	6.5 kW
Provedení	rack, 8U	rack, 6U





CEITEC

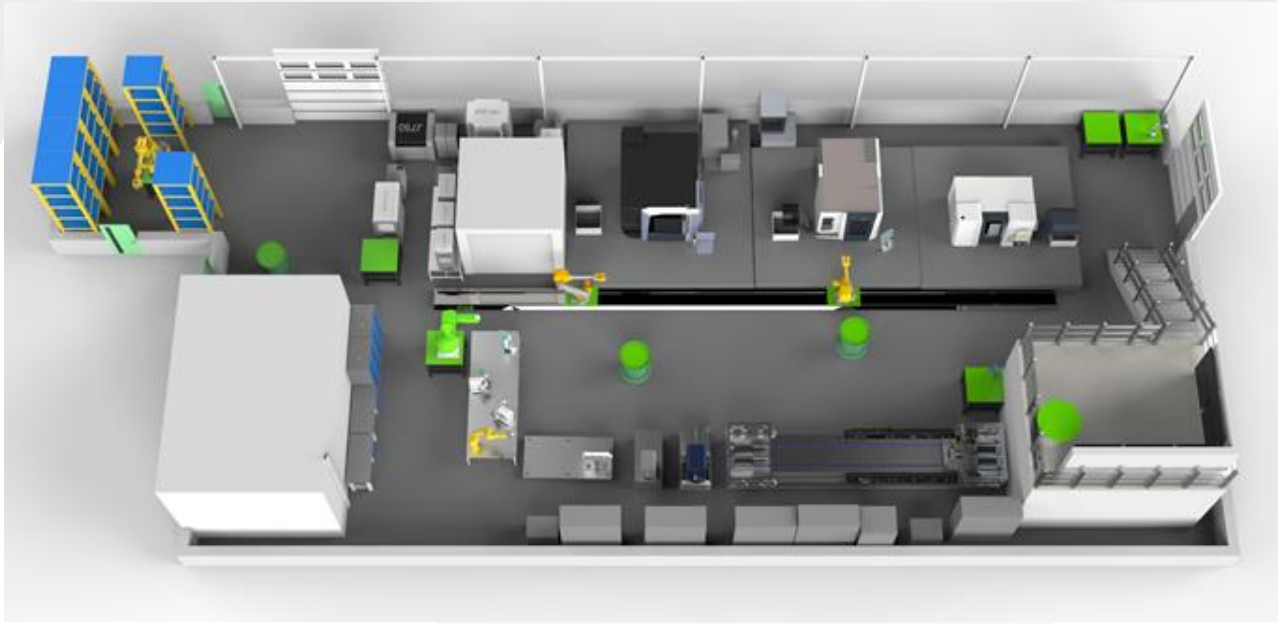
RICAIP TESTBED BRNO

Infrastruktura testbedu CEITEC VUT

Průmyslová hala – skutečné průmyslové prostředí

Vybavení

- 3D tisk
- Robotický sklad
- Laserové řezání/svařování
- 5osé a 3osé frézování, CNC soustruh
- Přesné 3D skenování
- Vybavení pro AR/VR
- Referenční optický lokalizační systém
- Automatizační HW, PLM SW
- Všesměrové mobilní roboty, průmyslové manipulátory, coboty
- Kráčející roboty (Boston Dynamics)
- Privátní 5G síť
- Dynamometry



RICAIP
TESTBED BRNO

CEITEC

RICAIP
Research and Innovation Centre
on Advanced Industrial Production

EDIH DIGIMAT

- **Evropský digitální inovační hub** podporující digitalizaci výrobních podniků
- Konsorcium - Intemac, VUT, JIC a Effectivity
- Zaměření na robotiku, automatizaci včetně využití AI
- Služby
  - **Test before invest**
  - **Experimenty na testbedu**
  - **Pomoc s nasazením digitálních technologií**
  - **Poradenství**
  - **Vzdělávací a popularizační aktivity (Digitální akademie)**
- Široké spektrum oblastí experimentů – robotika, 3D tisk, diagnostika strojů, aktuátory, UGV/AGV, AR/VR, lokalizace strojů/lidí, 5G komunikace, HPC,....

- Pro SME a small mid-caps (do 499 zaměstnanců) sleva 100%

TEF AI-MATTERS

The European Testing and Experimentation Facilities for AI in Manufacturing

- Síť evropských testbedů – Francie, Německo, Nizozemí, Česká republika, Španělsko, Dánsko, Itálie, Řecko
- TEF poskytuje služby pro celou EU
- Zaměření na nasazení AI ve výrobních firmách
  - **Experimenty s AI v průmyslovém prostředí**
  - **Vývoj ukázkových aplikací AI**
  - **Příprava datasetů**
  - **Podpora při vývoji aplikací AI pro výrobní technologie**
- Plánované významné další rozšíření technologií testbedu
- Poskytování služeb SME se slevou 100%

CEITEC
Financováno Evropskou unií NextGenerationEU
NÁRODNÍ PLÁN OBNOVY
Spolufinancováno Evropskou unií
AI matters

CEITEC Vysoké učení technické v Brně
Brno, Purkyňova 656/123
pavel.vaclavek@ceitec.vutbr.cz