Towards Efficient and Safe ML in Real-Time Cyber-Physical Systems

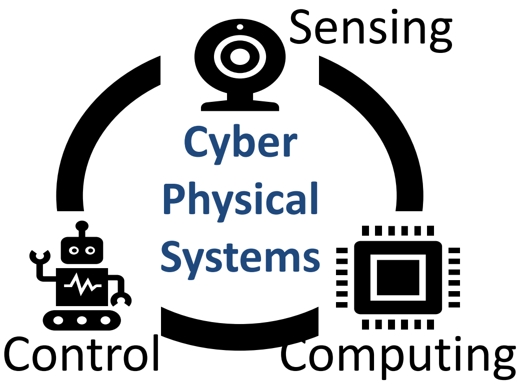
Zhishan Guo

Associate Professor, Computer Science, NC State University
zguo32@ncsu.edu

ML-RT-Agenda 2025 @Brussels

Cyber-Physical Systems (CPS)

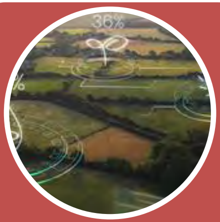
- Engineered systems that are built from, and depend upon, the seamless integration of computation and physical components.



- Real-Time Systems
  - *both* logical *and* temporal Correctness
  - Predictability: Provably guaranteed responses within strict time constraints (deadlines)

Efficient Computing & Automation | Real-Time Scheduling & Control | Wireless Networking

Wireless Sensing | Formal Methods | Constrain-Aware Security & Privacy

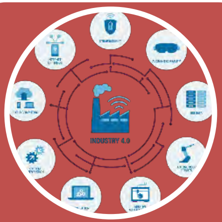


Agriculture

Environmental

Smart City

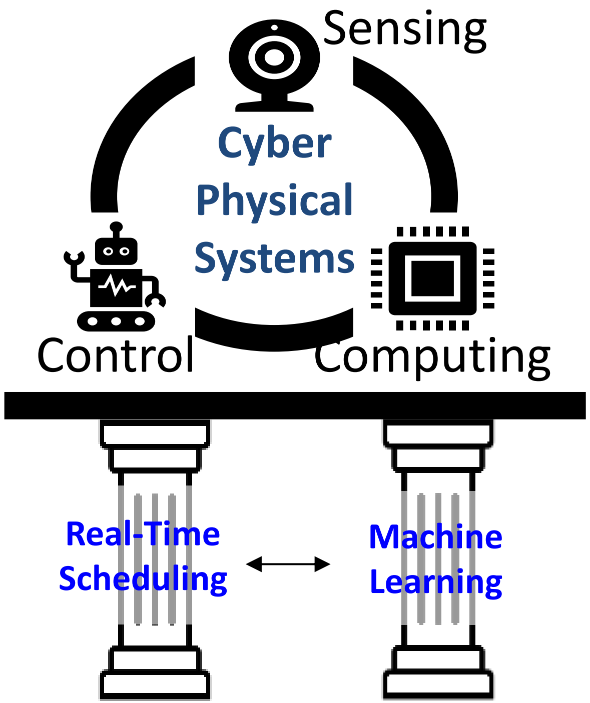
Healthcare



Robotics & Manufacturing

Autonomous Driving

Towards Efficient and Safe ML in Real-Time CPS



- Key Challenges in RT-CPS Design and Analysis
  - System Uncertainty & Reliability/Safety

    Application 1: F1-Tenth Autonomous Racing
  - Resource and Energy Efficiency

    Application 2: Cardiovascular Monitoring
  - Scalability and Curse of Dimensionality

    Application 3: Gait Kinematics Estimation

– System Uncertainty & Reliability/Safety

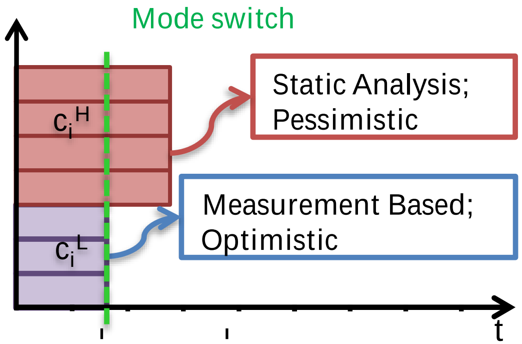
- ## Mixed Criticality Real-Time OS

[RTCSA'21] Reserving Processors by Precise Scheduling of Mixed-Criticality Tasks.

[RTNS'21] Precise Scheduling of Mixed-Criticality Tasks on Varying-Speed Multiprocessors.

[ICCAD'20] F2VD: Fluid Rates to VDs for Precise MC Scheduling on a Varying-Speed Processor.

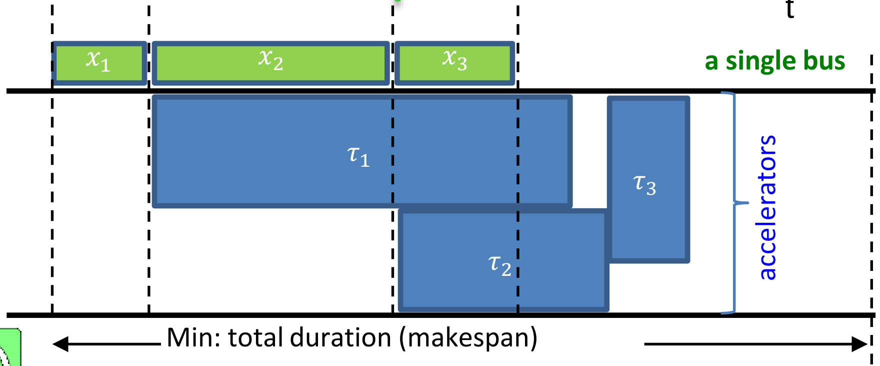
[RTSS'18] Precise Scheduling of Mixed-Criticality Tasks by Varying Processor Speed.



- ## CPU-GPU Co-Scheduling

[RTSS'19, TIME'21] MC scheduling of Gang tasks
(Outstanding paper award, Best student paper award)

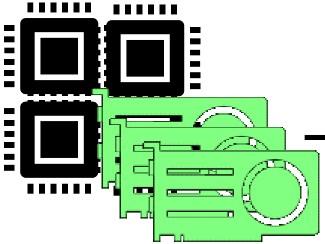
[RTAS 2020] Real-time scheduling upon a host-centric acceleration architecture with data offloading

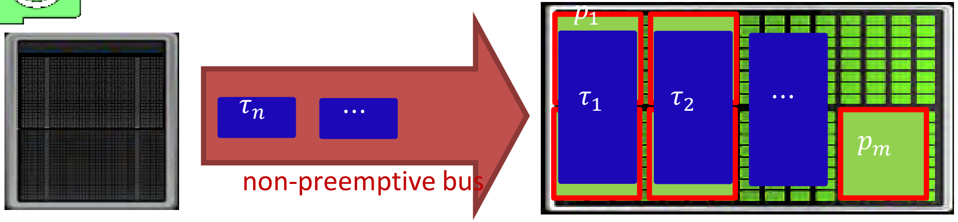


- ## IDK-Cascades

[RTS 2024] Scheduling IDK classifiers with arbitrary dependences to minimize the expected time to successful classification

[ECRTS'25] Optimizing IDK Cascades for Time-Series Input Streams





- Application 1: F1-Tenth Autonomous Racing

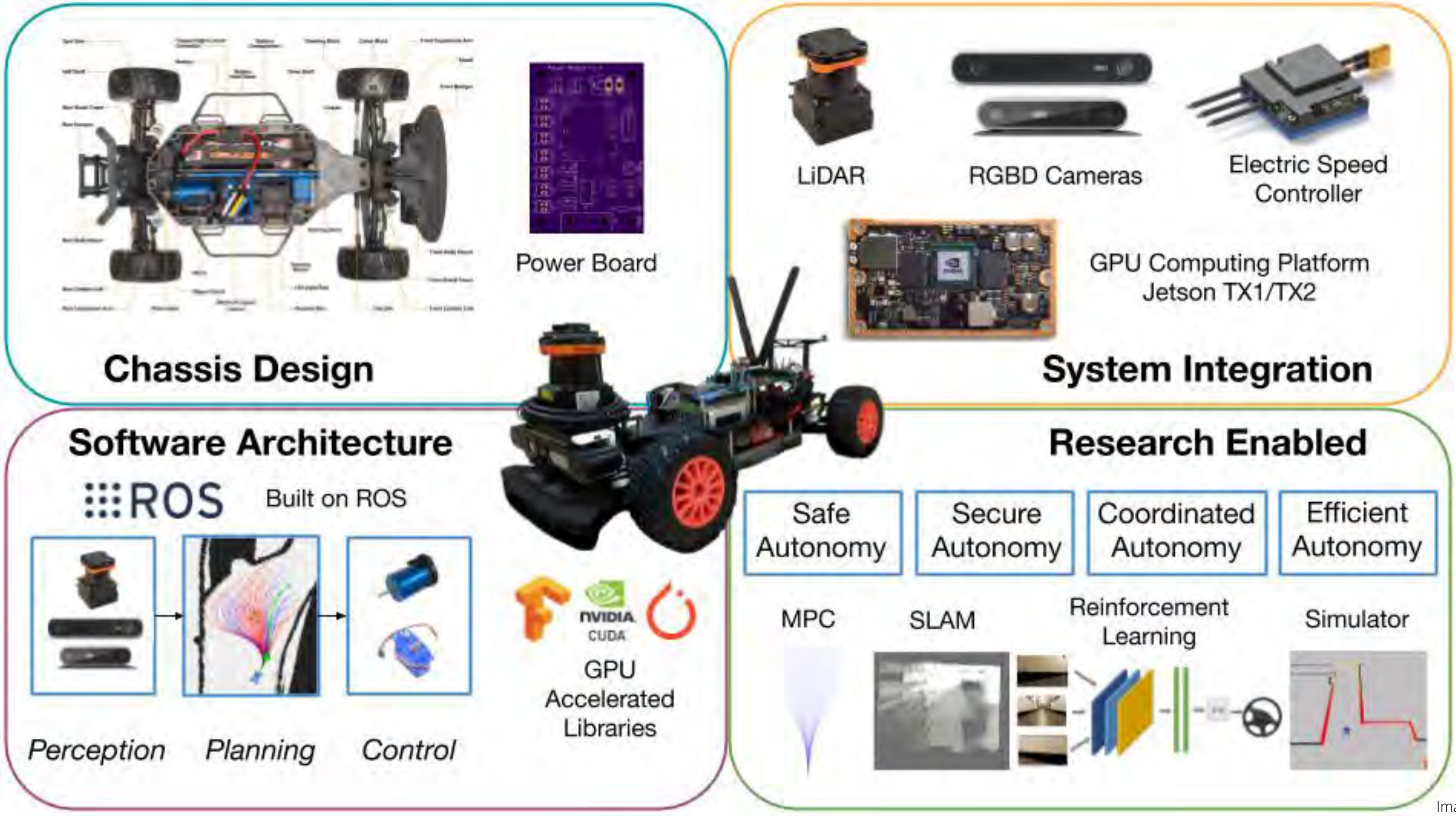


Image Source: linklab-uva

…mous Racing

ROS2 3-layer architecture: Software/Run-Time/RTOS

[RTAS'25] Physics-Informed MC Scheduling for F1Tenth Cars with Preemptable ROS 2 Executors

[HSCC'25] SOTERIA: A Formal Digital-Twin Framework for Safety-Assurance of Latency-Aware Safety-Critical CPS

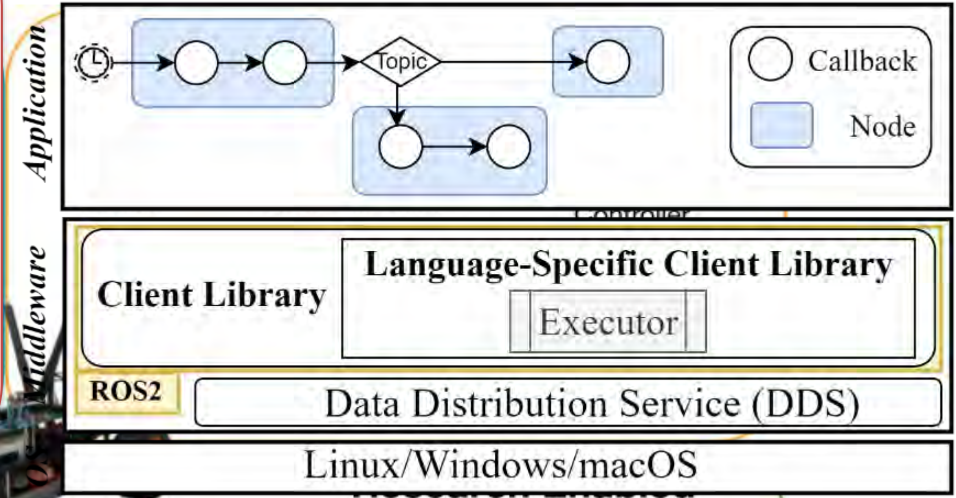
[TCAD 2024] Dynamic Priority Scheduling of Multithreaded ROS 2 Executor With Shared Resources

[RTSS'23] SEAM: An Optimal Message Synchronizer in ROS with Well-Bounded Time Disparity (Best Paper Award)

[RTAS'23] Real-Time Scheduling of Autonomous Driving System with Guaranteed Timing Correctness

[DAC'22] Response-Time Analysis for D'line-Based Scheduling of ROS2

[RTSS'21-Industry] Toward Real-Time Guaranteed Scheduling for Autonomous Driving Systems (Best Solution Award)



UPPAAL

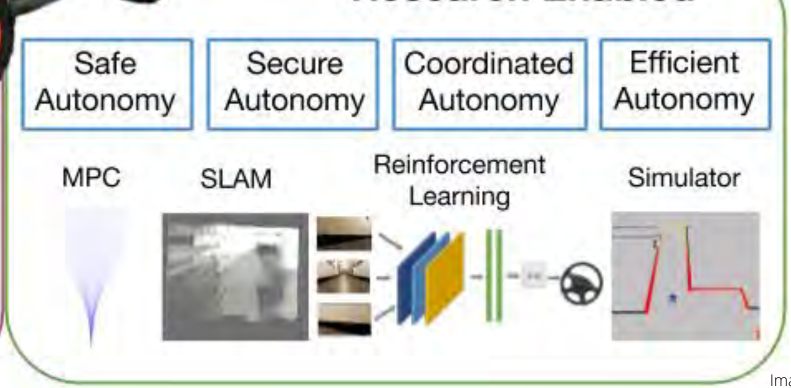
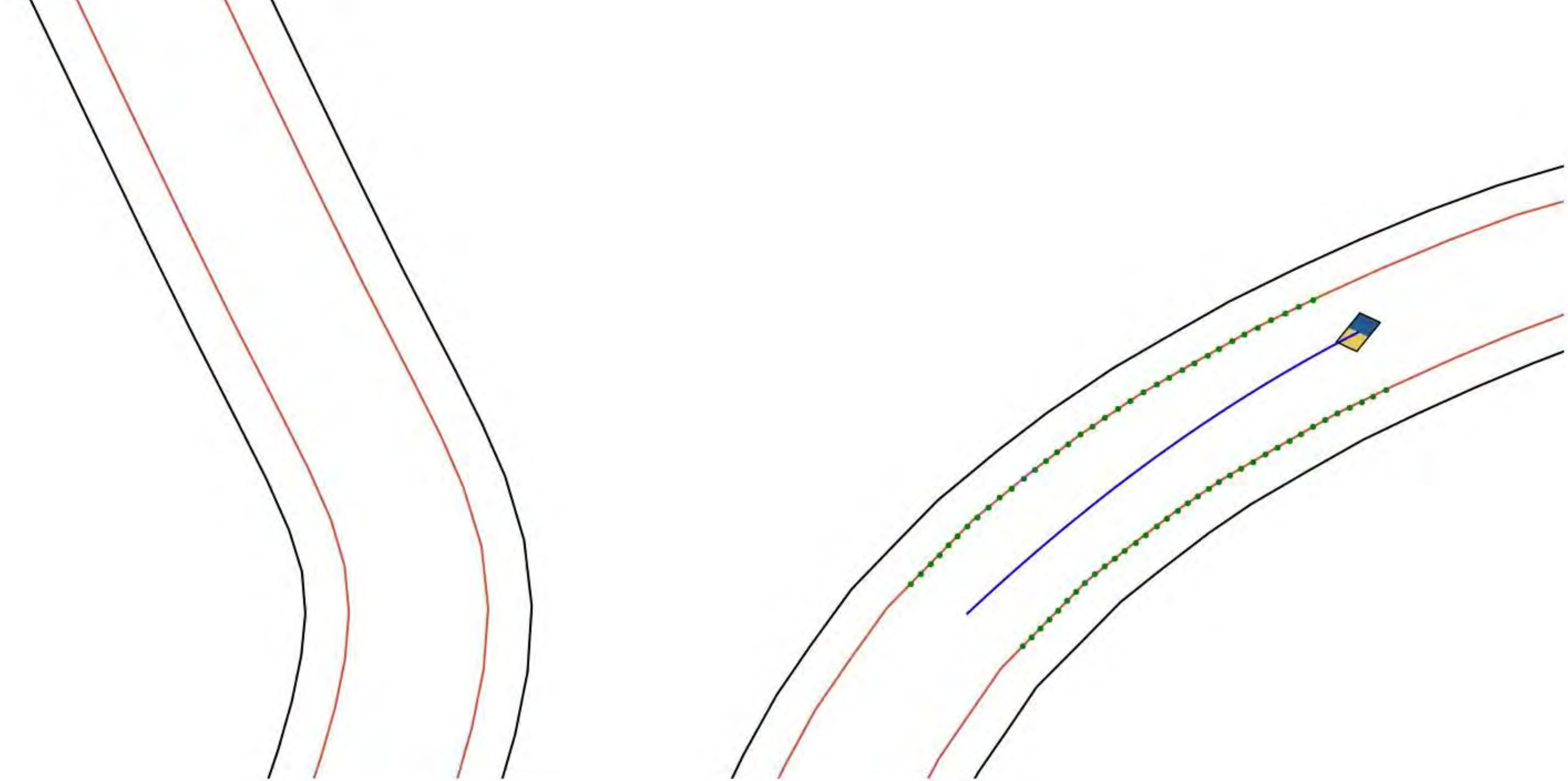


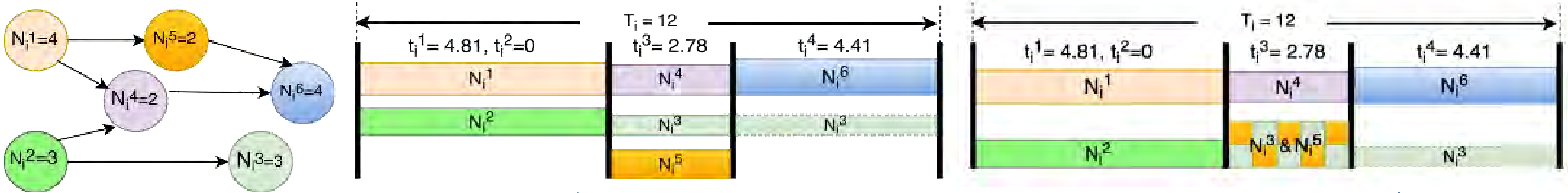
Image Source: linklab-uva

- Application 1: F1-Tenth Autonomous Racing



– Resource and Energy Efficiency

- DAG/Graph-Based Task



Decomposition and Frame Adjustment

Intra-Task Merging

Inter-Task/
Inter-Core Merging

- Energy-efficient scheduling of DAG tasks

[ECRTS'17, TECS'18] Multi-core processors

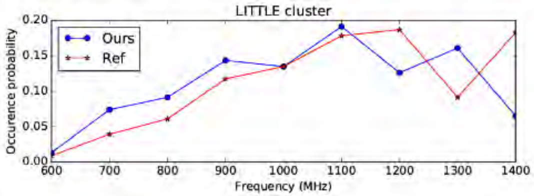
[RTAS'19, TPDS'20] Clustered multi-core processors

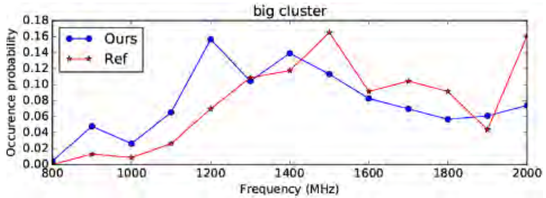
- Response time and feasibility analysis

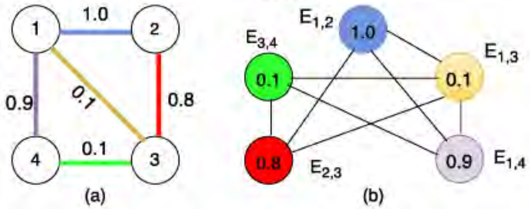
[RTSS'21] Calculating Worst-Case Response Time Bound for OpenMP Programs with Loop Structures.

[EMSOFT'20] Efficient Feasibility Analysis for Graph-based Real-Time Task Systems (Best paper award)

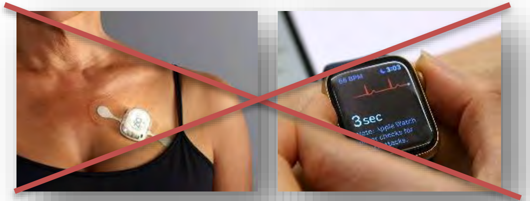
[DAC'20] On Computing Exact WCRT for DAG Task

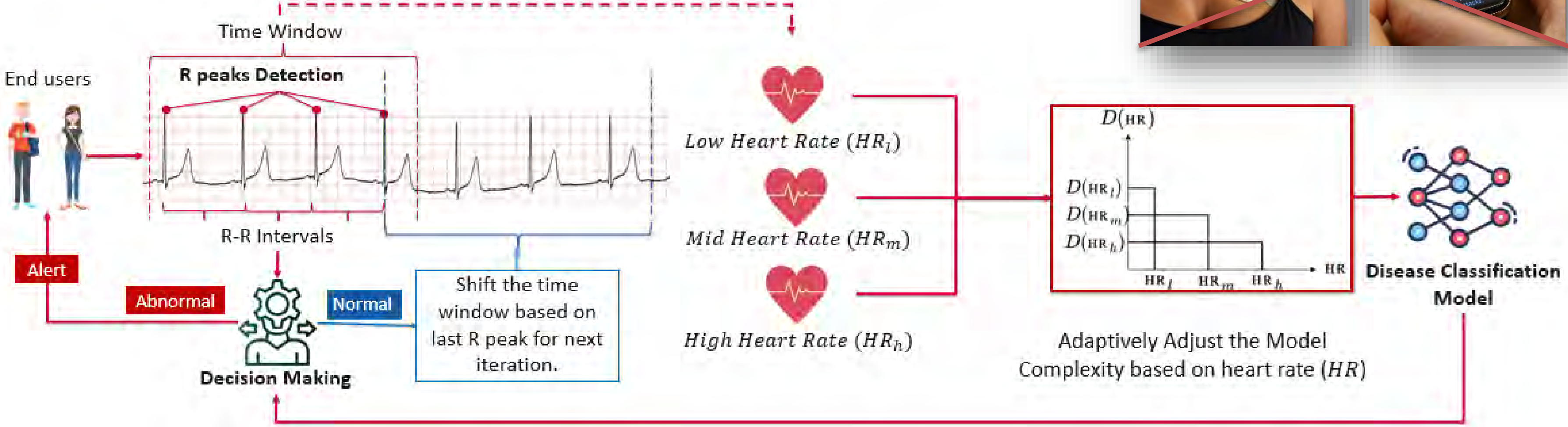






- Application 2: Real-time Cardiovascular Disease Detection on Embedded Devices





RQ1: How to balance the accuracy and inference speed, while preserving the memory usage?

[RTCSA'25] Adaptive Model Selection for Real-Time Cardiovascular Disease Detection on Embedded Systems
[ISNN'24] CardiacRT-NN: Real-Time Detection of Cardiovascular Disease Using Self-attention CNN-LSTM for Embedded Systems
[IPSN'24] Real-Time cardiovascular disease detection via abnormal electrocardiogram cycles on embedded systems
[ICCPS'15] Uniprocessor EDF scheduling of AVR task systems

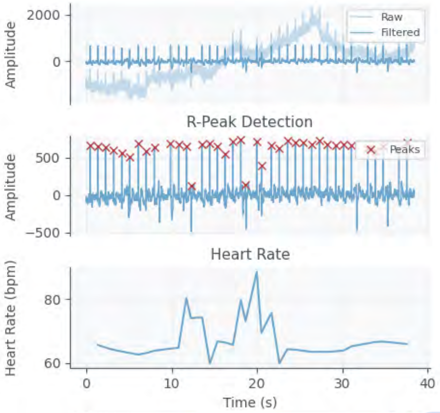
Heart Rate: Low

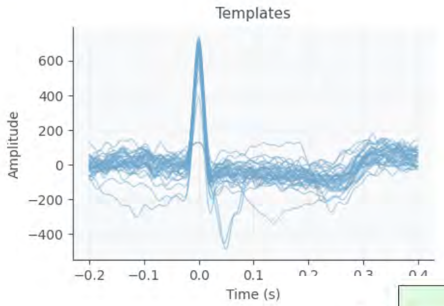
Heart Rate: Mid

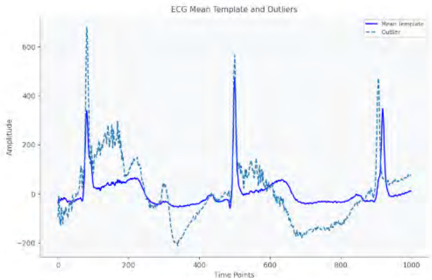
Heart Rate: High

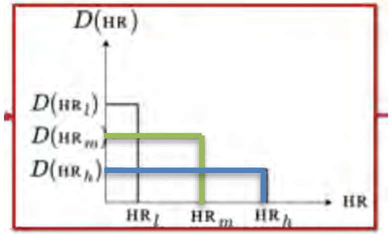
• Application 2: Real-time Cardiovascular Disease Detection

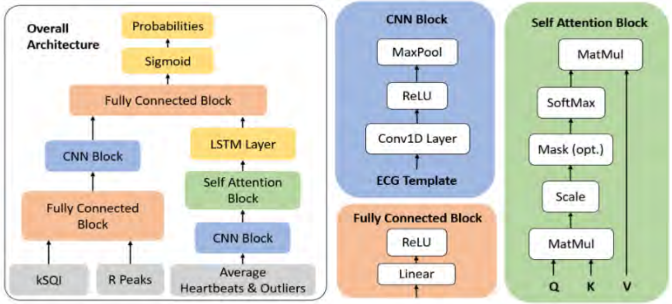
RQ2: How to design highly accurate DL models for abnormality detection?
RQ3: How to design a framework to handle varying heart rate conditions?

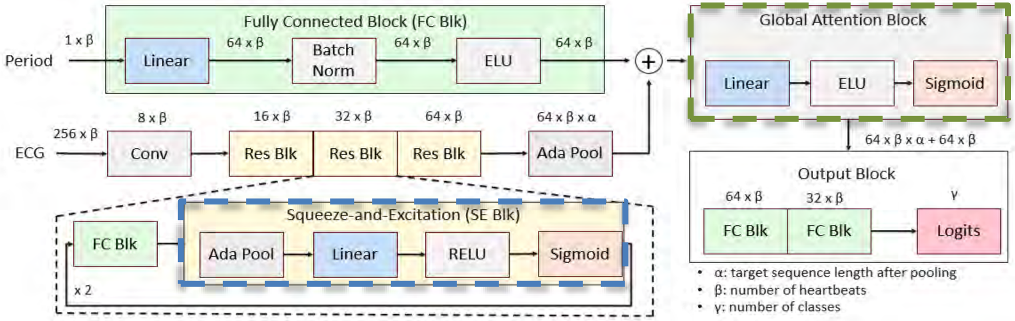












Self-Attention CNN-LSTM

– Scalability and Curse of Dimensionality

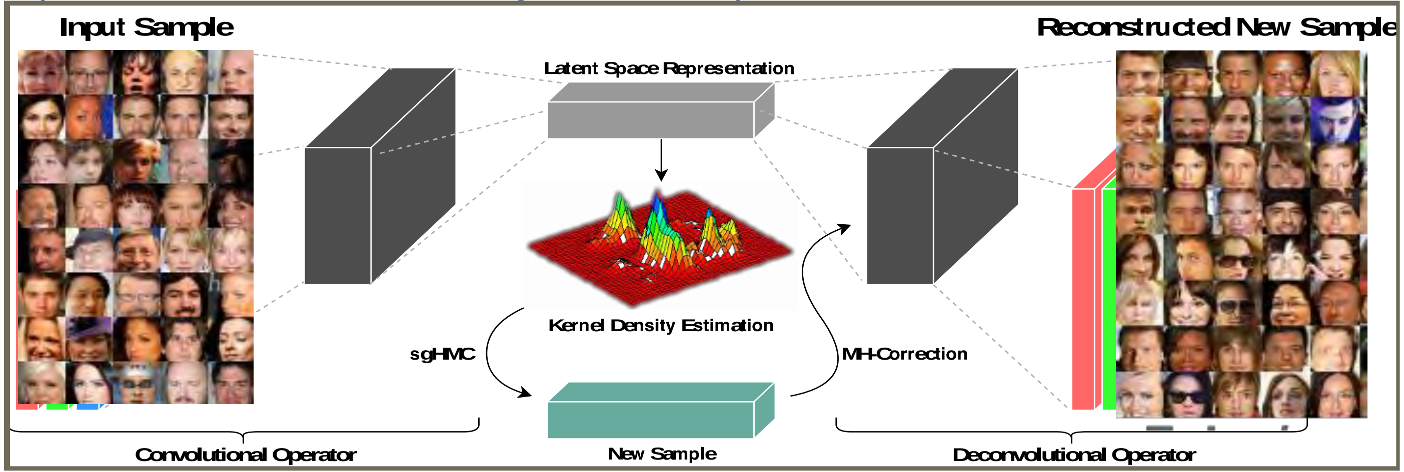
High Dimension and Low Sample Size (HDLSS) Settings

When the number of training samples is less than the dimension size of training sample, it can be treated as HDLSS settings.

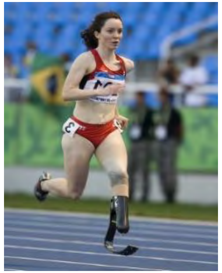
- LDA: inverse covariance matrix ??? (sparse regularization/ de-biased)
- Dataset Generation
  - Latent space explore for high quality sample generation
  - Estimated kernel density distribution
  - Explainable solution with strong control on space search

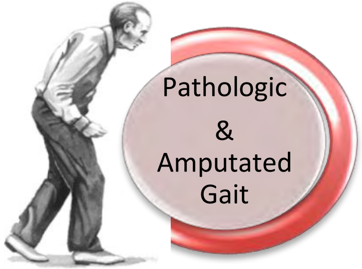
[ICDM'17] AWDA: An Adaptive Wishart Discriminant Analysis
[IJCAI'18] De-Biasing Covariance-Regularized LDA
[TNNLS'20] DBSDA: Lowering the Bound of Misclassification Rate for Sparse LDA via Model Debiasing

[AAAI'19] SpHMC: Spectral Hamiltonian Monte Carlo

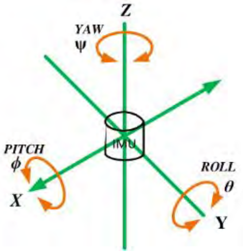


- Application 3: Gait Kinematics Estimation
  - Kinematics/joint angle
    - Disease progression
      Kinesiological health

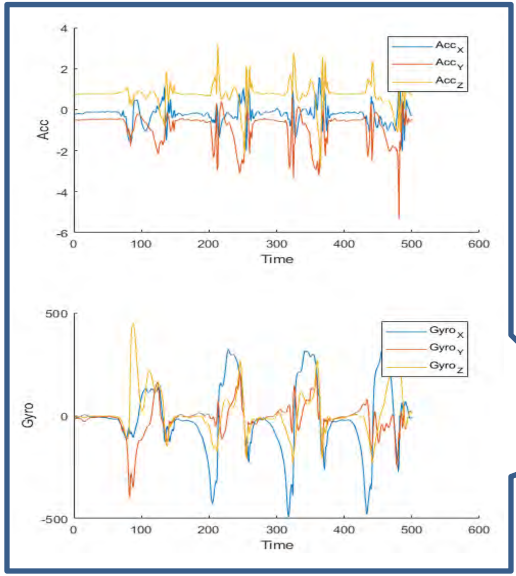


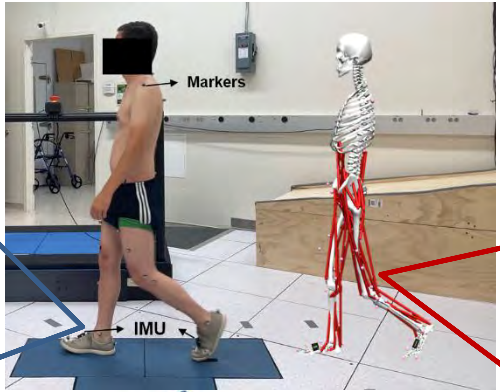


3 Axes
Accelerations

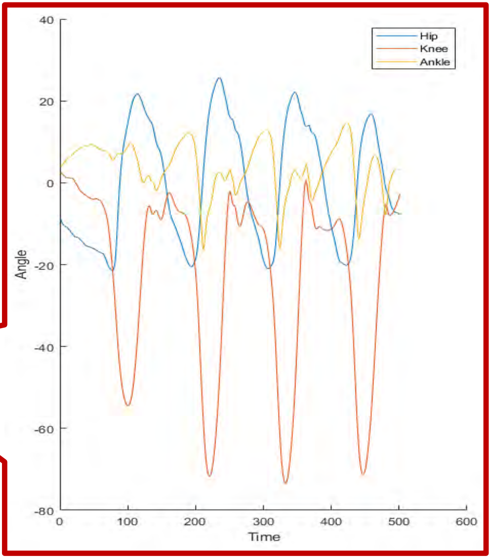


3 Axes
Angular Velocity

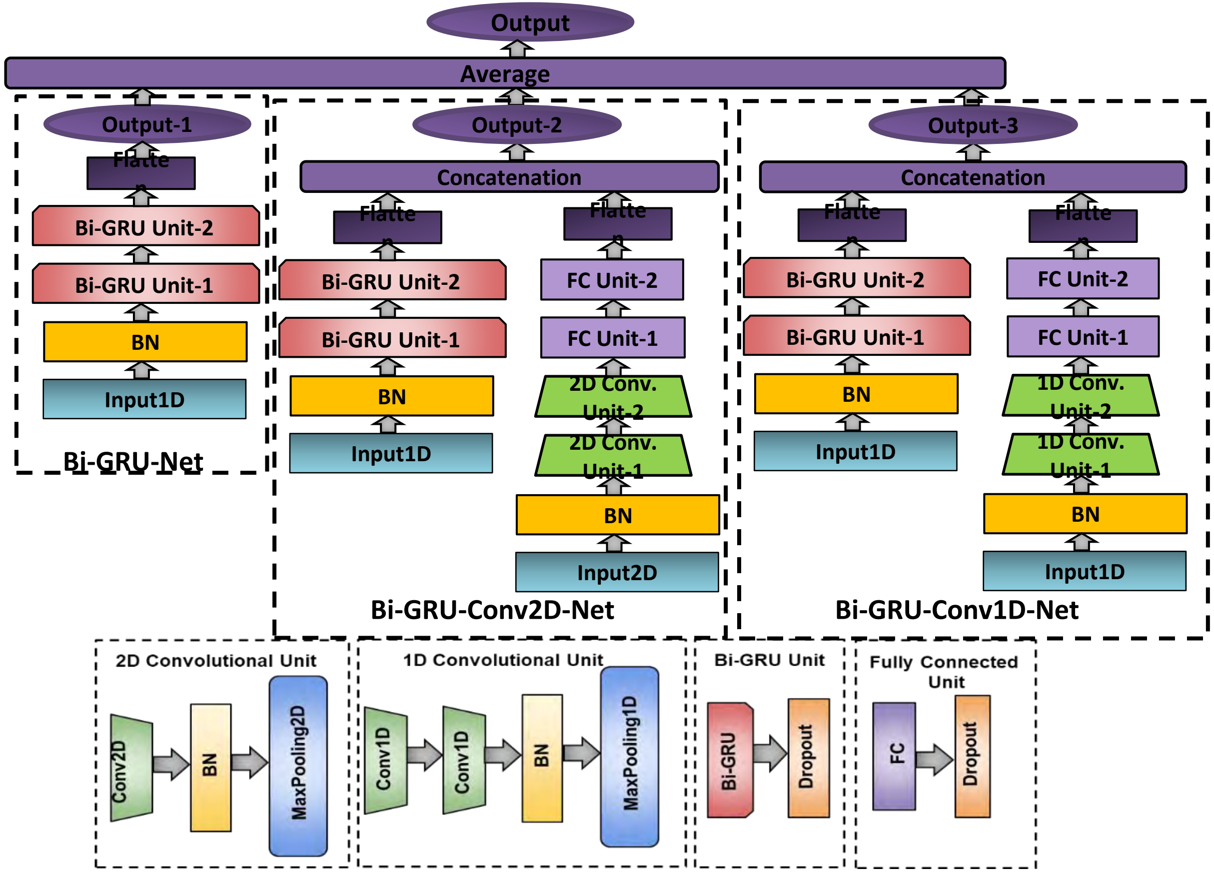


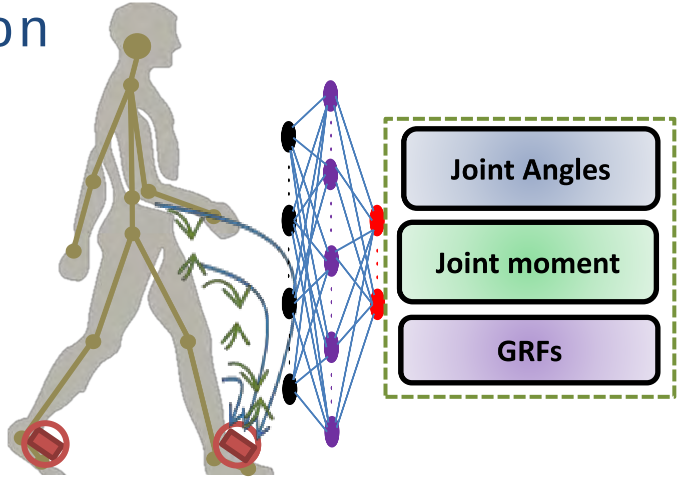


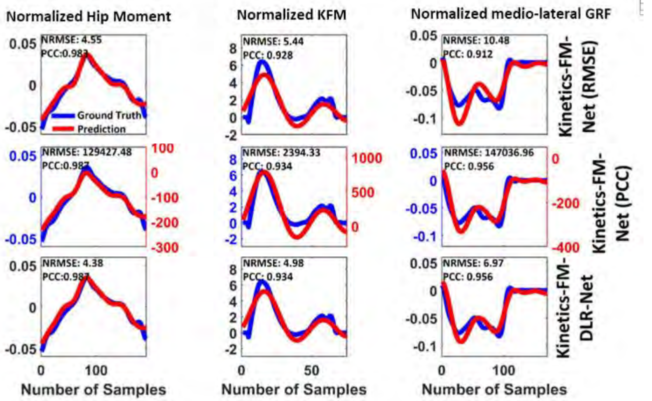
~ $10
Compact Size
Light Weight



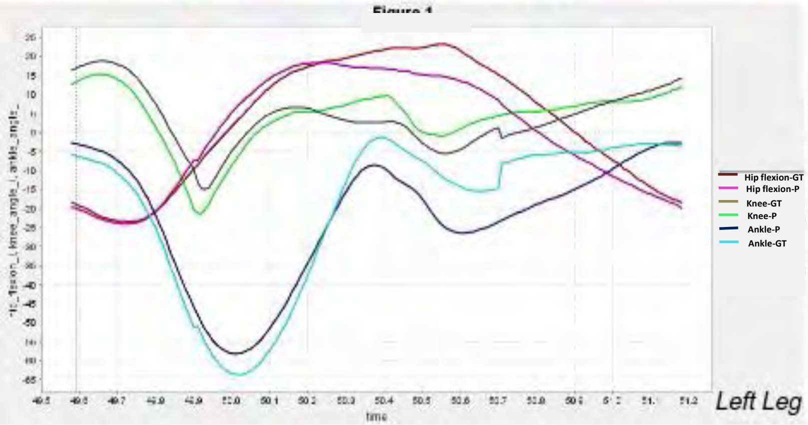
- Application 3: Gait Kinematics Estimation







- # Application 3: Gait Kinematics Estimation



[HICSS'24] Predicting Lower Extremity Joint Kinematics Using Multi-Modal Data in the Lab and Outdoor Environment
[JBHI2023] Wearable Motion Capture: Reconstructing and Predicting 3D Human Poses from Wearable Sensors
[JBHI2023] Estimation of Lower Extremity Joint Moments and 3D Ground Reaction Forces Using IMU Sensors in Multiple Walking Conditions: A Deep Learning approach
[JBHI2023] DeepBBWAE-Net: A CNN-RNN Based Deep Super Learner for Estimating Lower Extremity Sagittal Plane Joint Kinematics Using Shoe-Mounted IMU Sensors in Daily Living
[Chase'22] Estimation of Hip, Knee, and Ankle Joint Moment Using a Single IMU Sensor on Foot Via Deep Learning

Thank You!

PhD Students and Postdocs

Dr. Jiang Bian, Baidu Research
Prof. Ashik Ahmed Bhuiyan, West Chester U
Prof. Federico Reghenzani, Politecnico di Milano
Dr. Sudharsan Vaidhun, Cummins
Dr. Seyed Iman Taheri, University of Memphis
Dr. Ashkan Farhangi
Dr. Md Sanzid Bin Hossain
Abdullah Al Arafat
Kurt Wilson
Yixin (Eason) Li
Yuhan Zhao

Collaborators

Kunal Agrawal, WUSTL
Amro Awad, NC State U
John Baugh, NC State U
Sanjoy Baruah, WUSTL
Hwan Choi, MAE, UCF
Sajal Das, Missouri S&T
Anil Gehi, UNC-CH
Nan Guan, City U at HK
Xin Han, Dalian U Tech
Liting Hu, UCSC
Nan Hua, Rosen, UCF
Arthur Huang, Rosen, UCF
Denise Kay, Med., UCF
Jing Li, NJIT
Luca Santinelli, Airbus/BMW
Abu Saifullah, Iowa State U
Jinghao Sun, Dalian U Tech
Haoyi Xiong, Baidu Research
Wei Wang, UCLA
Jun Wang, ECE, UCF
Don Wunsch, Missouri S&T
Kecheng Yang, Texas State U

Sponsors

CYBER FLORIDA
NSF
NIH
NHK
Meta
UNC SCHOOL OF MEDICINE North Carolina Translational and Clinical Sciences Institute