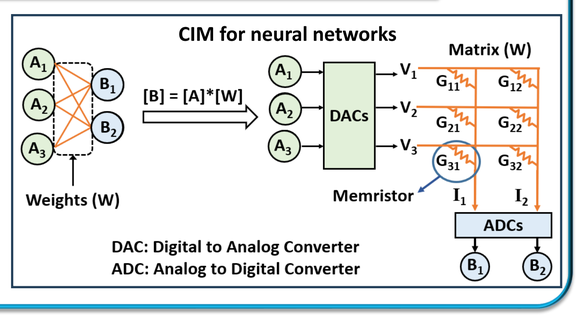
Computation-In-Memory based Edge-AI for Healthcare: A Cross-Layer Approach

Dr. Sumit Diware, *Delft University of Technology, The Netherlands*

DATE-2025 PhD Forum

1. Motivation

- Edge computing for AI (edge-AI)
  - Better latency, privacy, and security
- Computation-in-memory (CIM)
  - Energy efficient edge-AI
  - Critical domain: Healthcare

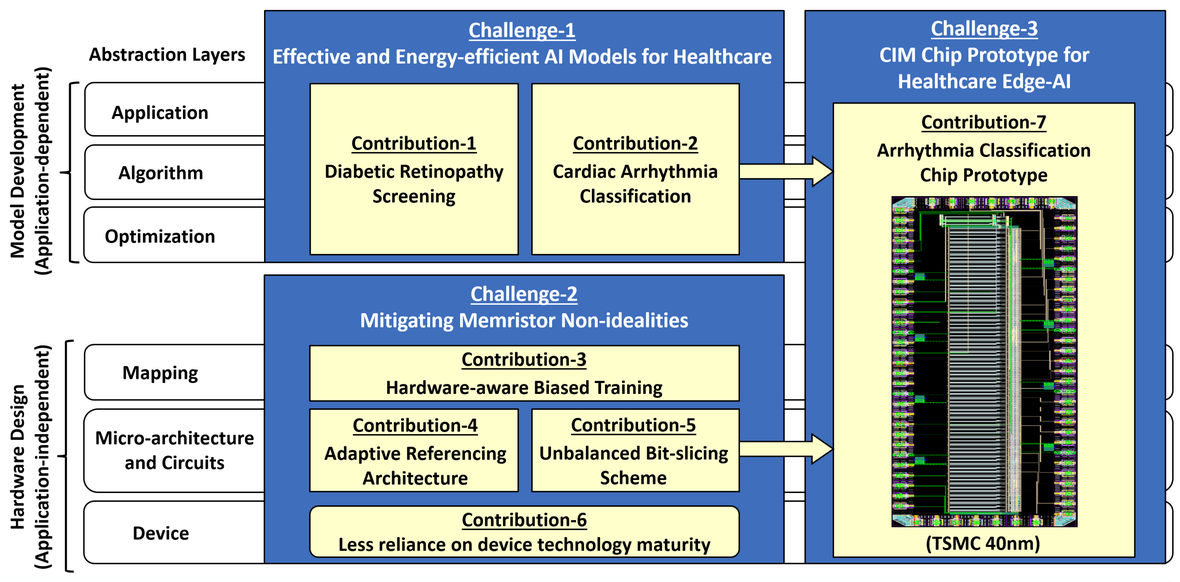


2. Challenges

- Challenge-1: Effective & energy-efficient healthcare AI models
  - Existing works ignore energy-efficiency & practical effectiveness
- Challenge-2: Mitigating memristor non-idealities
  - Existing non-ideality mitigation schemes are ineffective
- Challenge-3: CIM chip prototypes for healthcare
  - Lack of CIM AI chips providing non-ideality mitigation & effective healthcare

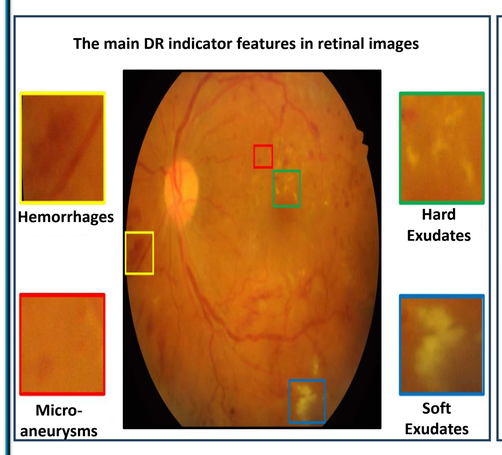
3. Thesis Contributions Overview

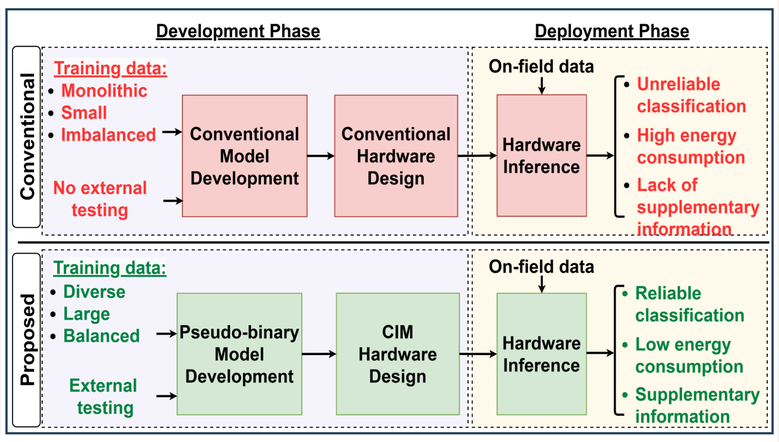
- Holistic cross-layer research covering the entire abstraction stack



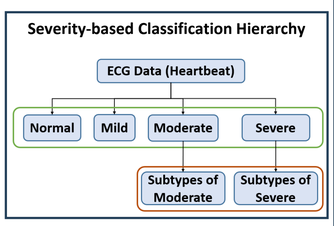
4. Contributions for Challenge-1

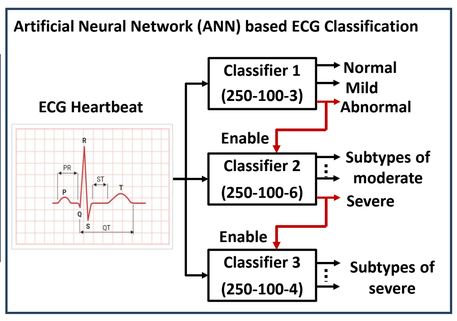
- Diabetic Retinopathy (DR) Screening [1,2]
  - Early detection of DR (leading cause of blindness worldwide)
  - Reliable & accurate DR screening at three orders of magnitude less energy

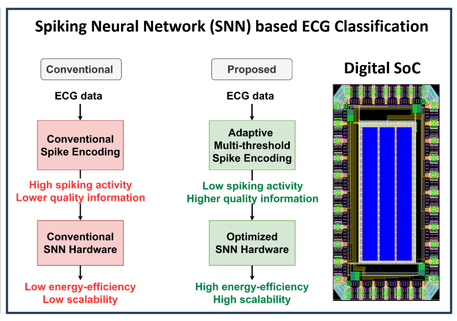




- Cardiac Arrhythmia Classification [3,4]
  - Early diagnosis of cardiovascular diseases (leading cause of global deaths)
  - High accuracy with energy improvement of 25×(as ANN) to 50×(as SNN)

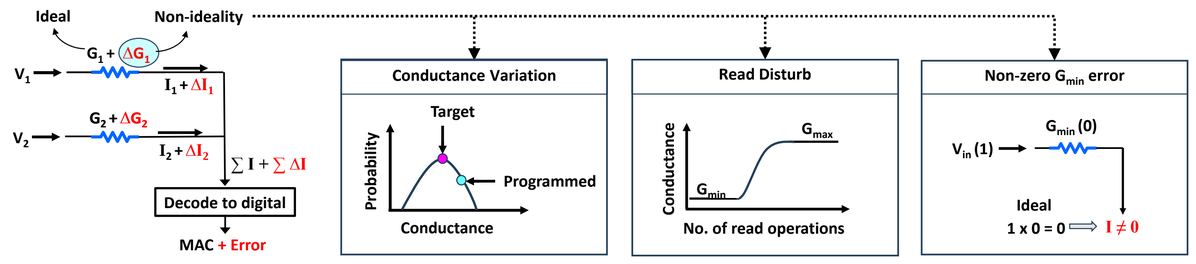




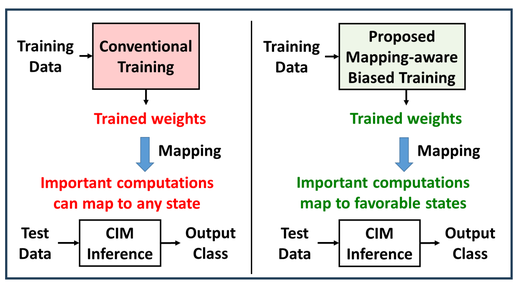


5. Contributions for Challenge-2

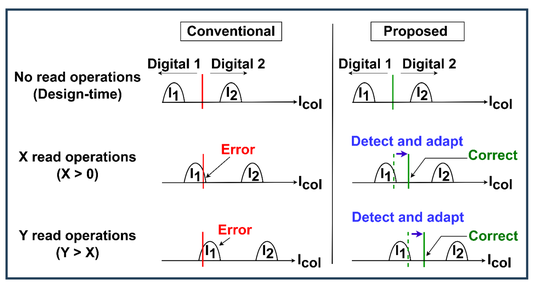
- Memristor non-idealities



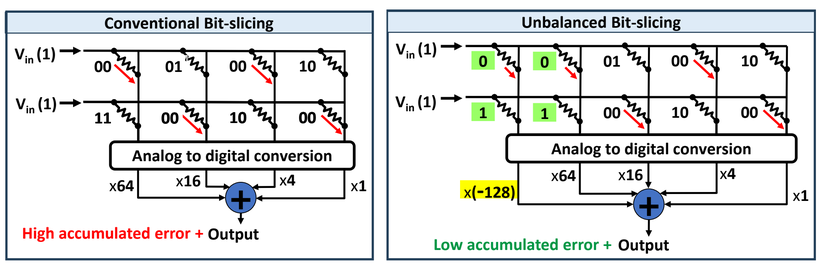
- Hardware-aware Biased Training Algorithm [5]
  - Mitigates conductance variation
  - 2.4× hardware accuracy

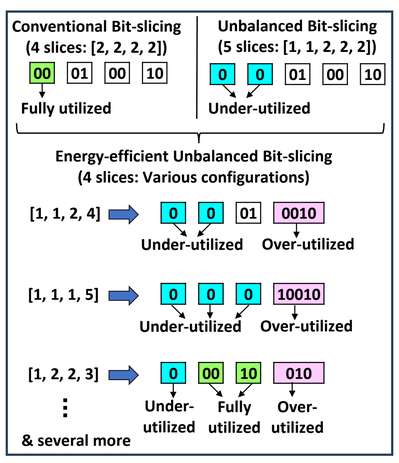


- Adaptive Referencing Architecture [6]
  - Mitigates read disturb
  - 2× hardware accuracy



- Unbalanced Bit-slicing [7,8]
  - Mitigates non-zero G_{min} error
  - 7.3× hardware accuracy

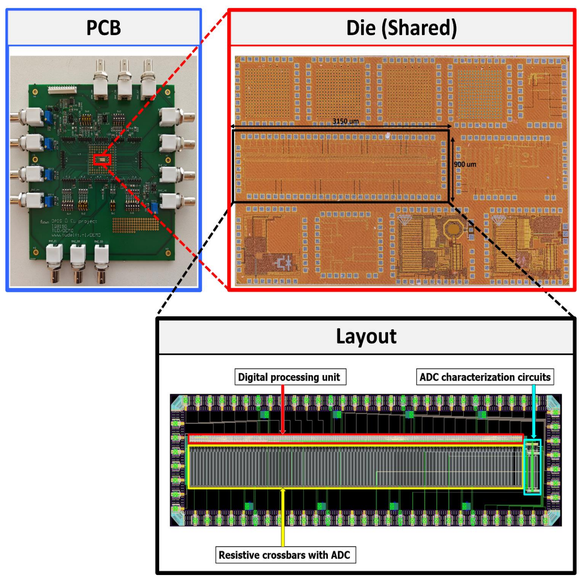




- Less reliance on device technology maturity
  - Showed high hardware accuracy despite the non-idealities

6. Contribution for Challenge-3

- CIM System-on-Chip Prototype for Arrhythmia Classification [9]
  - Integrates our ECG classifier with non-ideality mitigations
  - Optimized data sampling and hardware-aware quantization
  - TSMC 40nm technology
  - 2.9 sq. mm. Si area
  - 100 MHz clock frequency
  - CIM crossbars with novel ADC
  - Optimized digital processing unit
  - ADC characterization provisions



New Publication Ongoing

7. Conclusions

- The proposed models advance the effectiveness of AI-powered healthcare and its edge deployment
- The proposed mitigation schemes enable CIM edge AI product development without waiting for ideal memristor devices
- Integrations of our models and non-ideality mitigation schemes are well-suited for wearable/handheld healthcare devices

TU Delft

31 March - 2 April 2025 | Lyon, France

DATE 25 Design, Automation and Test in Europe Conference | The European Event for Electronic System Design & Test

Publications:

1. S. Diware et al., 2025, Springer Nature
2. S. Diware et al., 2024, IEEE Access
3. S. Diware et al., 2023,TBioCAS
4. S. Diware et al., 2025, DATE
5. S. Diware et al., 2023, AICAS
6. S. Diware et al., 2024, AICAS
7. S. Diware et al., 2022, TETCI
8. S. Diware et al., 2021, AICAS
9. S. Diware et al., 2024, ICCAD