tinyML® Talks

Enabling Ultra-low Power Machine Learning at the Edge

"Analysis of ECG Data by Energy Efficient Decision Trees on a Reconfigurable ASIC"

Tobias Peikenkamp - Hahn-Schickard

July 7, 2021

TINY ML

www.tinyML.org

tinyML Talks Sponsors

arm
tinyML Strategic Partner

Deeplite

EDGE IMPULSE
tinyML Strategic Partner

emza visual sense
tinyML Strategic Partner

GREENWAVES TECHNOLOGIES
tinyML Strategic Partner

LatentAI
Adaptive AI for a Smarter Edge
tinyML Strategic Partner

HOTG
tinyML Strategic Partner

maxim integrated™

Qeexo

Qualcomm
tinyML Strategic Partner

Reality AI®
tinyML Strategic Partner

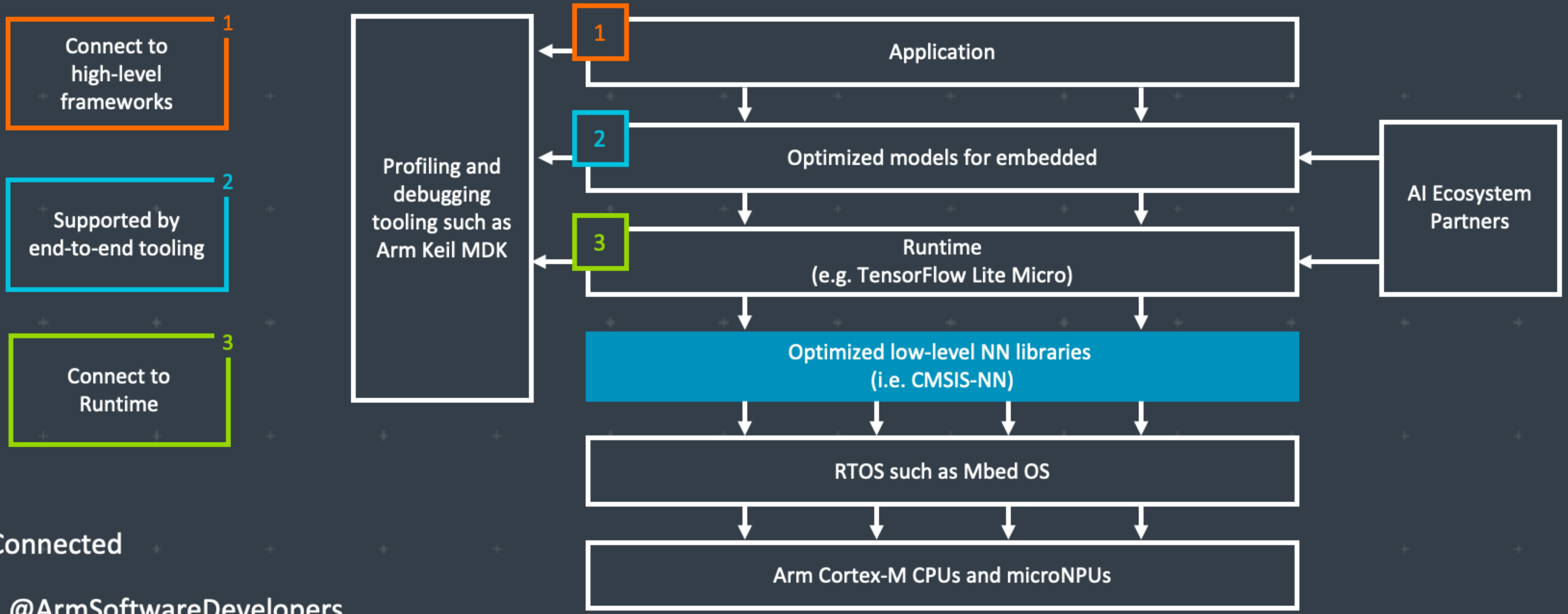
SensiML™
tinyML Strategic Partner

SynSense

SYNTIANT
tinyML Strategic Partner

Additional Sponsorships available – contact Olga@tinyML.org for info

Arm: The Software and Hardware Foundation for tinyML



Stay Connected

@ArmSoftwareDevelopers

@ArmSoftwareDev

Resources: developer.arm.com/solutions/machine-learning-on-arm

Deeplite

WE USE AI TO MAKE OTHER AI FASTER, SMALLER AND MORE POWER EFFICIENT

Automatically compress SOTA models like MobileNet to <200KB with **little to no drop in accuracy** for inference on resource-limited MCUs

Reduce model optimization trial & error from weeks to days using Deeplite's **design space exploration**

APPLY NOW

Deploy more models to your device without sacrificing performance or battery life with our **easy-to-use software**

BECOME BETA USER **bit.ly/testdeeplite**

mobilityXlab arm AI 100 CBINSIGHTS

TinyML for all developers

C++ library

Arduino library

WebAssembly

Dataset

Acquire valuable training data securely

Enrich data and train ML algorithms

Edge Device

Real sensors in real time

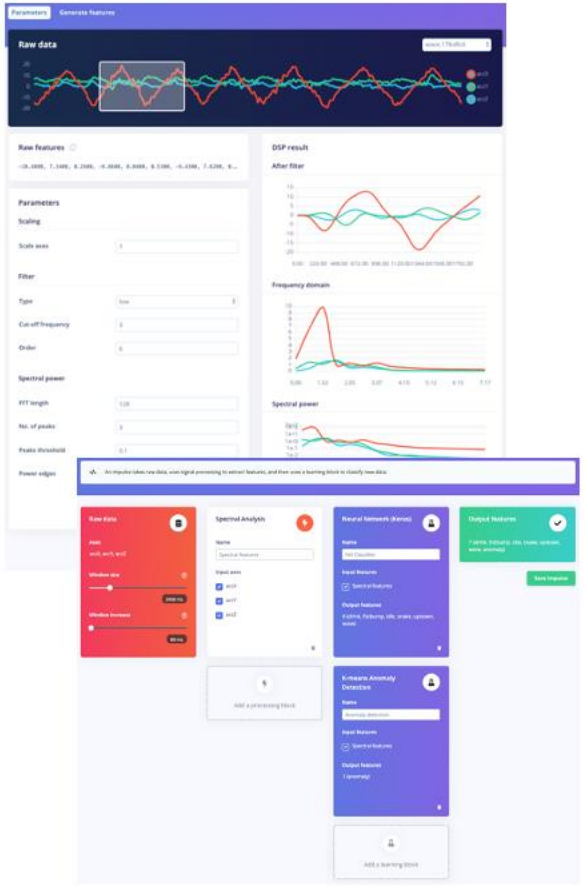
Open source SDK

Impulse

Embedded and edge compute deployment options

Test impulse with real-time device data flows

Test



www.edgeimpulse.com

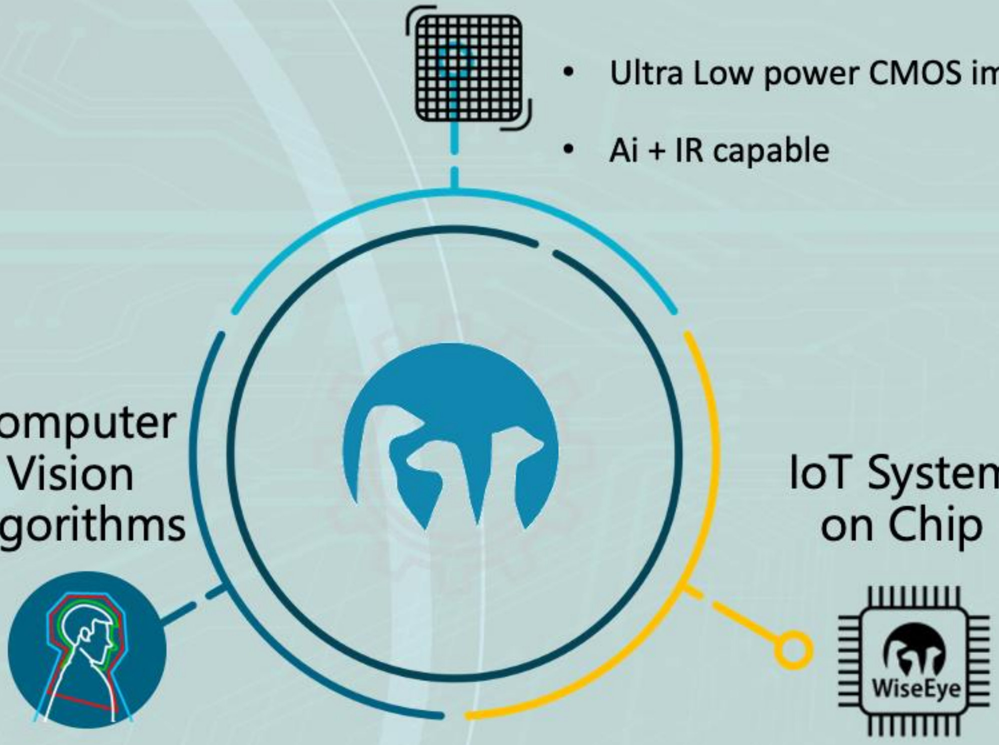
emza
visual sense

The Eye in IoT

Edge AI Visual Sensors

CMOS Imaging Sensor

- Ultra Low power CMOS imager
- Ai + IR capable



Computer Vision Algorithms

- Machine Learning algorithm
- <1MB memory footprint
- Microcontrollers computing power
- Trained algorithm
- Processing of low-res images
- Human detection and other classifiers

IoT System on Chip

- Machine Learning edge computing silicon
- <1mW always-on power consumption
- Computer Vision hardware accelerators

info@emza-vs.com

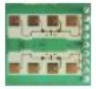
Enabling the next generation of Sensor and Hearable products to process rich data with energy efficiency

Visible Image

Sound

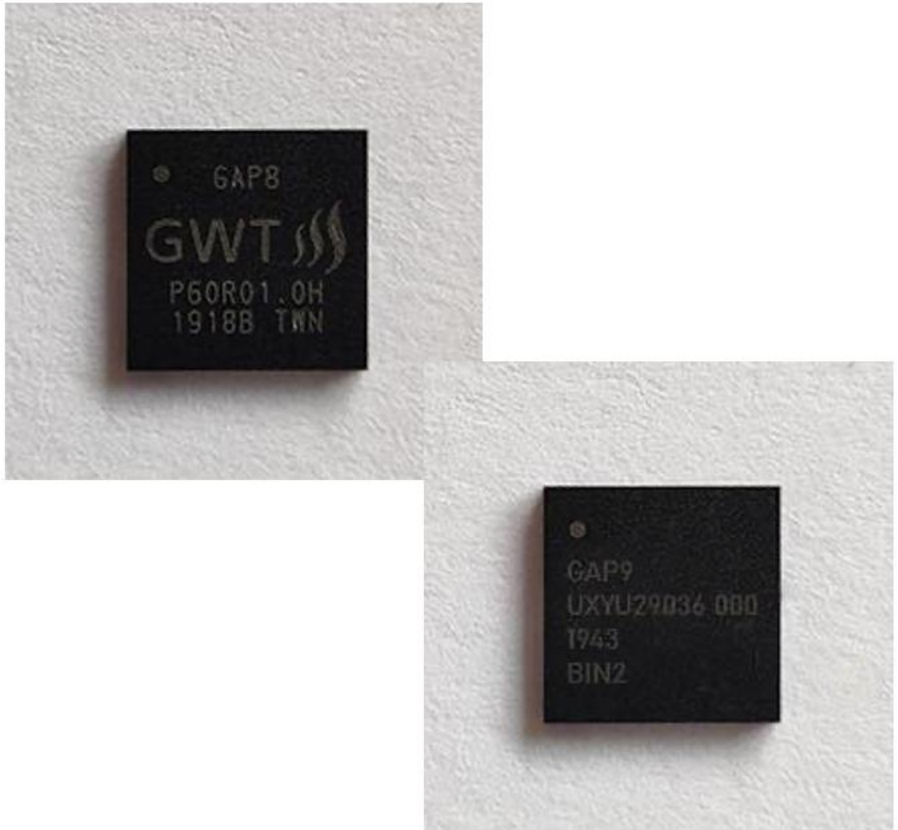
IR Image

Radar



Bio-sensor

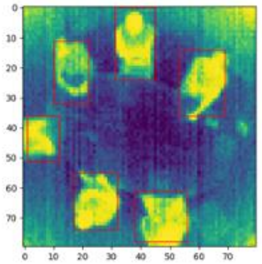
Gyro/Accel



Wearables / Hearables

Battery-powered consumer electronics

IoT Sensors



GREENWAVES TECHNOLOGIES

LatentAI

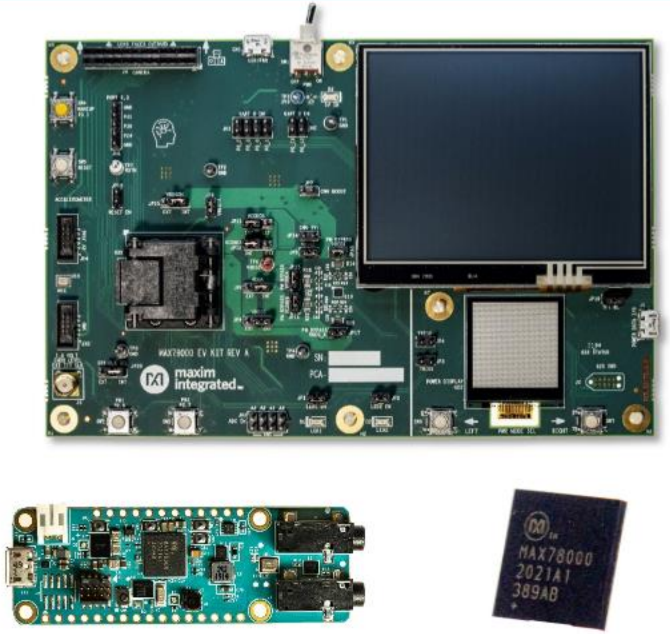
Adaptive AI for the Intelligent Edge

Latentai.com

HOTG

Maxim Integrated: Enabling Edge Intelligence

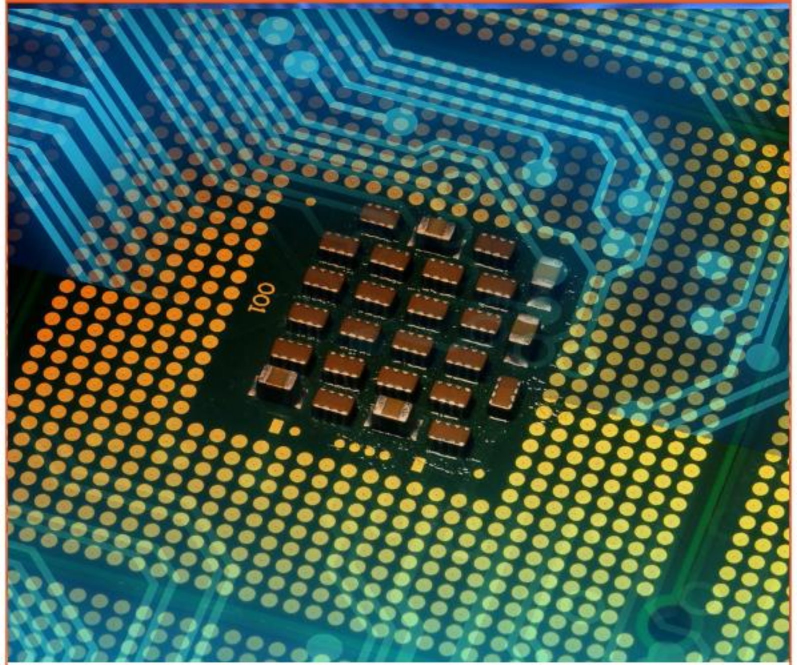
Advanced AI Acceleration IC



The new MAX78000 implements AI inferences at low energy levels, enabling complex audio and video inferencing to run on small batteries. Now the edge can see and hear like never before.

www.maximintegrated.com/MAX78000

Low Power Cortex M4 Micros



Large (3MB flash + 1MB SRAM) and small (256KB flash + 96KB SRAM, 1.6mm x 1.6mm) Cortex M4 microcontrollers enable algorithms and neural networks to run at wearable power levels.

www.maximintegrated.com/microcontrollers

Sensors and Signal Conditioning

Health sensors measure PPG and ECG signals critical to understanding vital signs. Signal chain products enable measuring even the most sensitive signals.

www.maximintegrated.com/sensors

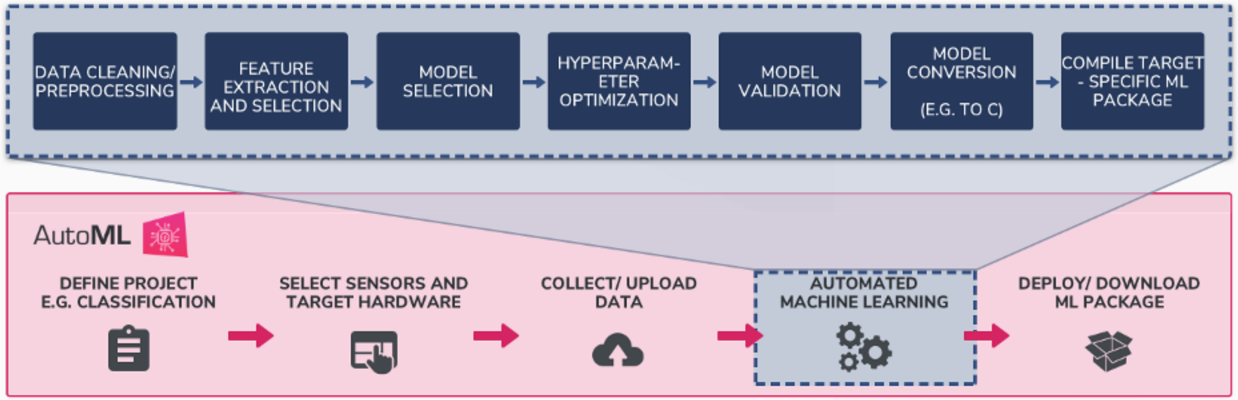
Qeexo AutoML

Automated Machine Learning Platform that builds tinyML solutions for the Edge using sensor data

Key Features

- Supports 17 ML methods:
  - Multi-class algorithms: GBM, XGBoost, Random Forest, Logistic Regression, Gaussian Naive Bayes, Decision Tree, Polynomial SVM, RBF SVM, SVM, CNN, RNN, CRNN, ANN
  - Single-class algorithms: Local Outlier Factor, One Class SVM, One Class Random Forest, Isolation Forest
- Labels, records, validates, and visualizes time-series sensor data
- On-device inference optimized for low latency, low power consumption, and small memory footprint applications
- Supports Arm® Cortex™- M0 to M4 class MCUs

End-to-End Machine Learning Platform



For more information, visit: www.qeexo.com

Target Markets/Applications

- Industrial Predictive Maintenance
- Smart Home
- Wearables
- Automotive
- Mobile
- IoT

Qualcomm
AI research

Advancing AI research to make efficient AI ubiquitous

Power efficiency

Model design, compression, quantization, algorithms, efficient hardware, software tool

Personalization

Continuous learning, contextual, always-on, privacy-preserved, distributed learning

Efficient learning

Robust learning through minimal data, unsupervised learning, on-device learning

A platform to scale AI across the industry

Perception

Object detection, speech recognition, contextual fusion

Reasoning

Scene understanding, language understanding, behavior prediction

Action

Reinforcement learning for decision making

Edge cloud

Cloud

IoT/IIoT

Automotive

Mobile

Reality AI®

Add Advanced Sensing to your Product with Edge AI / TinyML

https://reality.ai info@reality.ai @SensorAI Reality AI

Pre-built Edge AI sensing modules, plus tools to build your own

Reality AI solutions

- Prebuilt sound recognition models for indoor and outdoor use cases
- Solution for industrial anomaly detection
- Pre-built automotive solution that lets cars "see with sound"

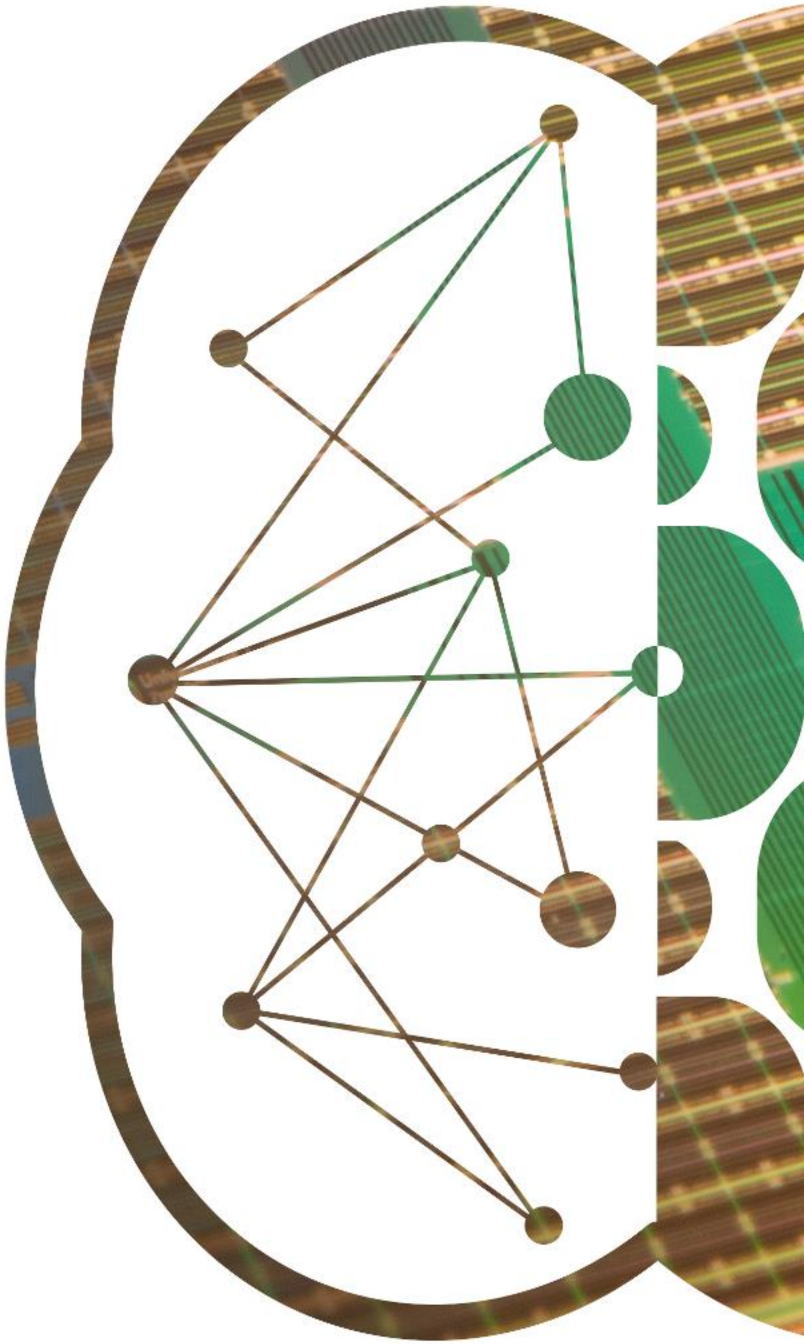
Reality AI Tools® software

- Build prototypes, then turn them into real products
- Explain ML models and relate the function to the physics
- Optimize the hardware, including sensor selection and placement

SynSense

SynSense builds **sensing and inference** hardware for **ultra-low-power** (sub-mW) **embedded, mobile and edge** devices. We design systems for **real-time always-on smart sensing**, for audio, vision, IMUs, bio-signals and more.

https://SynSense.ai



SYNTIANT

Syntiant Corp. is moving artificial intelligence and machine learning from the cloud to edge devices. Syntiant's chip solutions merge deep learning with semiconductor design to produce ultra-low-power, high performance, deep neural network processors. These network processors enable always-on applications in battery-powered devices, such as smartphones, smart speakers, earbuds, hearing aids, and laptops. Syntiant's Neural Decision Processors™ offer wake word, command word, and event detection in a chip for always-on voice and sensor applications.

Founded in 2017 and headquartered in Irvine, California, the company is backed by Amazon, Applied Materials, Atlantic Bridge Capital, Bosch, Intel Capital, Microsoft, Motorola, and others. Syntiant was recently named a CES® 2021 Best of Innovation Awards Honoree, shipped over 10M units worldwide, and unveiled the NDP120 part of the NDP10x family of inference engines for low-power applications.

www.syntiant.com

@Syntiantcorp

collaboration with

Focus on:
(i) developing new use cases/apps for tinyML vision; and (ii) promoting tinyML tech & companies in the developer community

Submissions accepted until August 20th, 2021
Winners announced on September 1, 2021 ($6k value)
Sponsorships available: *sponsorships@tinyML.org*

https://www.hackster.io/contests/tinyml-vision

Next tinyML Talks

Date	Presenter	Topic / Title
Tuesday, July 13	**Siddharth Tallur,** Associate Professor, Electrical Engineering, IIT Bombay	Edge-compatible machine learning algorithms for vibration condition based monitoring of machines

Webcast start time is 8 am Pacific time

Please contact talks@tinyml.org if you are interested in presenting

Local Committee in Germany

Alexis Veynachter,
Master Degree in Control Engineering, Senior Field Application Engineer
Infineon 32bits MCUs for Sensors, Fusion & Control

Carlos Hernandez-Vaquero
Software Project Manager, IoT devices
Robert Bosch

Prof. Dr. Daniel Mueller-Gritschneder
Interim Head - Chair of Real-time Computer Systems
Group Leader ESL - Chair of Electronic Design Automation
Technical University of Munich

Marcus Rüb
Researcher in the field of TinyML
Hahn-Schickard

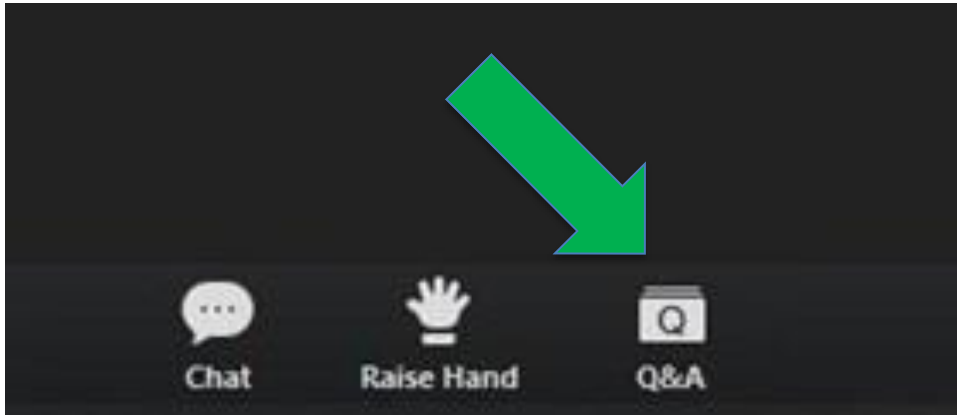
Reminders

Slides & Videos will be posted tomorrow

tinyml.org/forums

youtube.com/tinyml

Please use the Q&A window for your questions



Tobias Peikenkamp

Tobias Peikenkamp studied computer science at the Carl von Ossietzky University Oldenburg with the focus on autonomous systems, machine learning and Artificial Intelligence. After completing the Master of Science in 2019, he started working at the AI department of Hahn-Schickard where his work involved analysis of sensor data and the application of decision tree ensembling techniques.

Hahn Schickard

Visions to Products

Analysis of ECG Data by Energy Efficient Decision Trees on a Reconfigurable ASIC

Tobias Peikenkamp

7.7.2021

Hahn-Schickard

Hahn-Schickard-Gesellschaft für angewandte Forschung e.V.

Hahn Schickard

39 M€ revenue (2020)

- 9.5 M€ - contracts with industry
- 6.9 M€ - institutional funding BW

256 employees (2020)

Member of the Innovation Alliance Baden-Württemberg

innBW INNOVATIONSALLIANZ BADEN-WÜRTTEMBERG

Institute of Micro Assembly Technology
Stuttgart (ISO 9001:2015)

Institute of Micro and Information Technology
Villingen-Schwenningen (ISO 9001:2015)

Institute of Microanalysis Systems
Freiburg (ISO 13485:2015)

Institute of Microanalysis Systems
Ulm

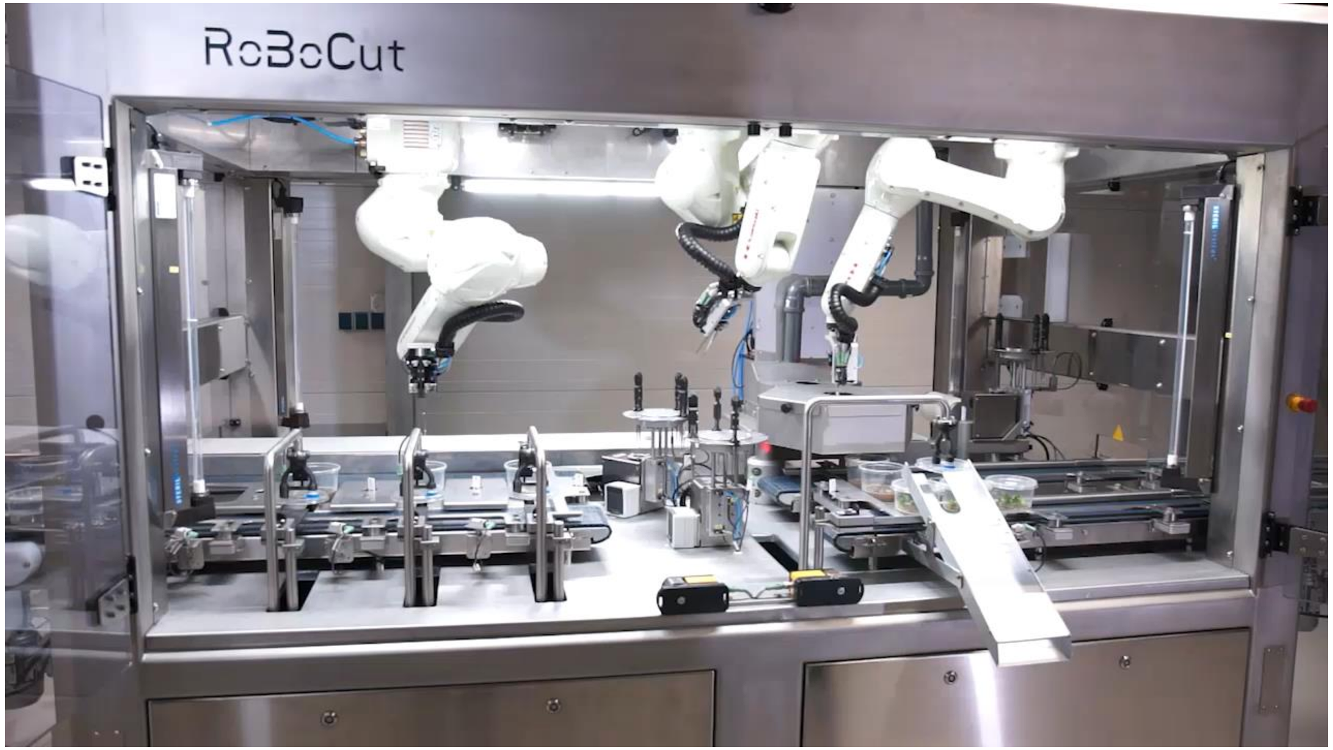
Artificial Intelligence (AI)

Focus

- Embedded AI
- AI and privacy

Projects

- Industrial Robot Interaction with AI (FMER)
- Self learning AI radio location system
- Skin cancer diagnosis
- AI Hardware accelerator
- AI-Trainer



Are you a German SME and interested in AI?

Hahn Schickard

Free AI offers for German SMEs

Mittelstand 4.0 Kompetenzzentrum Textil vernetzt

KI MITTELSTAND

Bundesministerium für Wirtschaft und Energie

Our offer for small and medium-sized enterprises (SMEs):

- On-site visits to develop suitable fields of application for AI
- AI trainings
- AI events
- AI development projects for your application

AiF Mitglied innBW INNOVATIONSALLIANZ BADEN-WÜRTTEMBERG

Are you a German SME and interested in AI?

For small and medium-sized enterprises (SMEs) in the Schwarzwald-Baar-Heuberg district:

- AI Working Groups & Expert Tables
- AI Information Series
- AI Co-Working Spaces & Prototyping Trainings
- AI Consulting

Dates:

- 08.07.21 KI for Predictive Maintenance

Are you a SME from Baden-Württemberg and interested in TinyML?

Within the scope of QuickChecks, we analyse whether and how your application-related problem can be supported by the use of Edge AI.

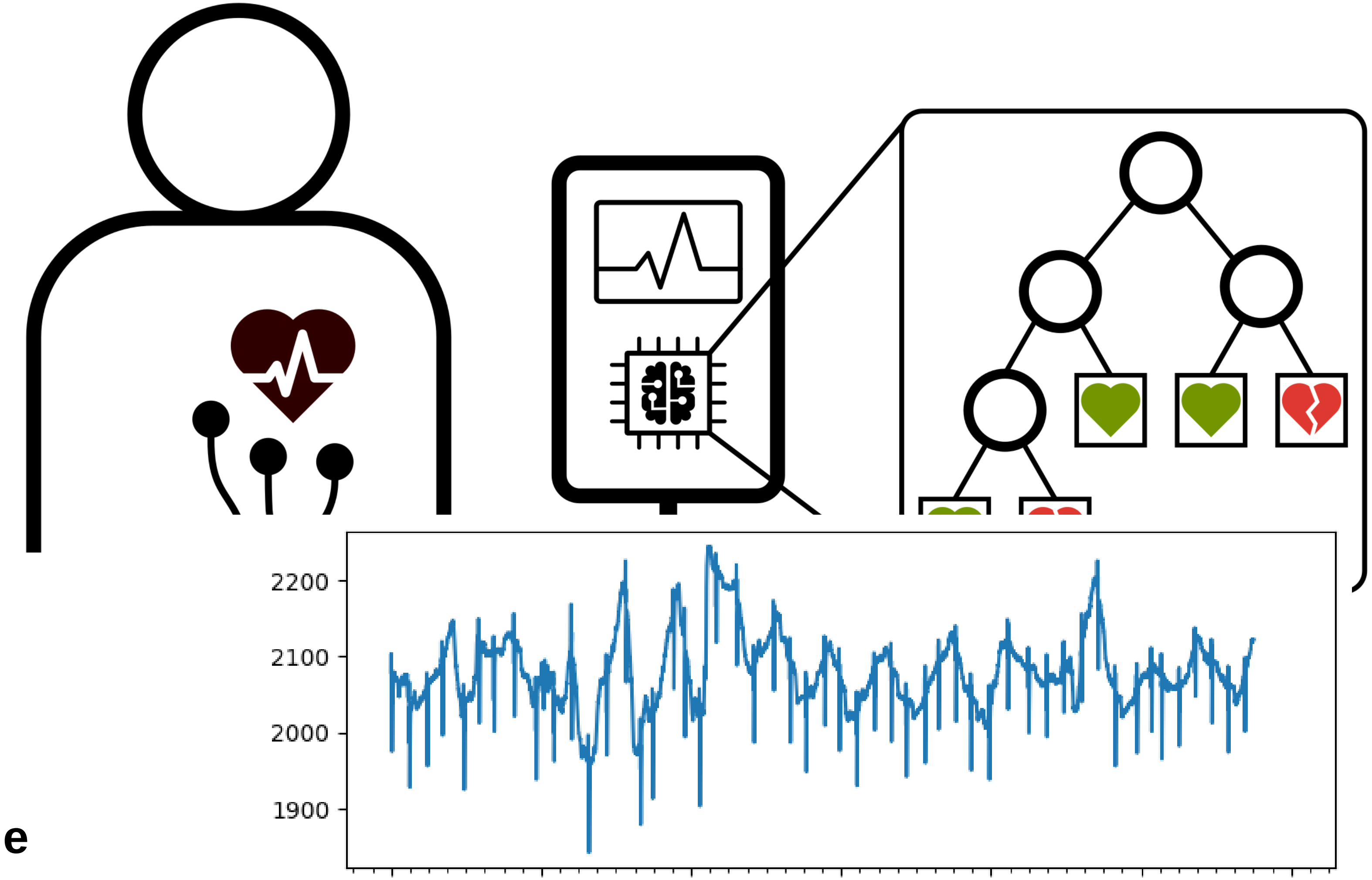
We help you to outline the solution path, analyse your current status and help you select the resources and accompany you during the first technical implementation.

Apply now for a free QuickCheck -> quickcheck@ims-chips.de

Pilote Innovation Initiative "KI-Sprung"

Contest held by FMER

- Classification on atrial fibrillation
  - Energy efficiency
  - Accuracy
  - "Innovativeness"
- 500 data sets
  - 2 minutes of ECG data
  - 512 Hz
  - Clean data
    - Low noise / disturbance
    - Distinct classes
- Participation in 22nm category
- End December 2020
- Win awarded in March 2021



Concept Decision Tree Ensemble

Metrics & Results

Contest criteria

- Min. Recall 90%
- Max. Fallout 20%
- Minimum energy for classification

⇨ Focus on recognition of sick patients

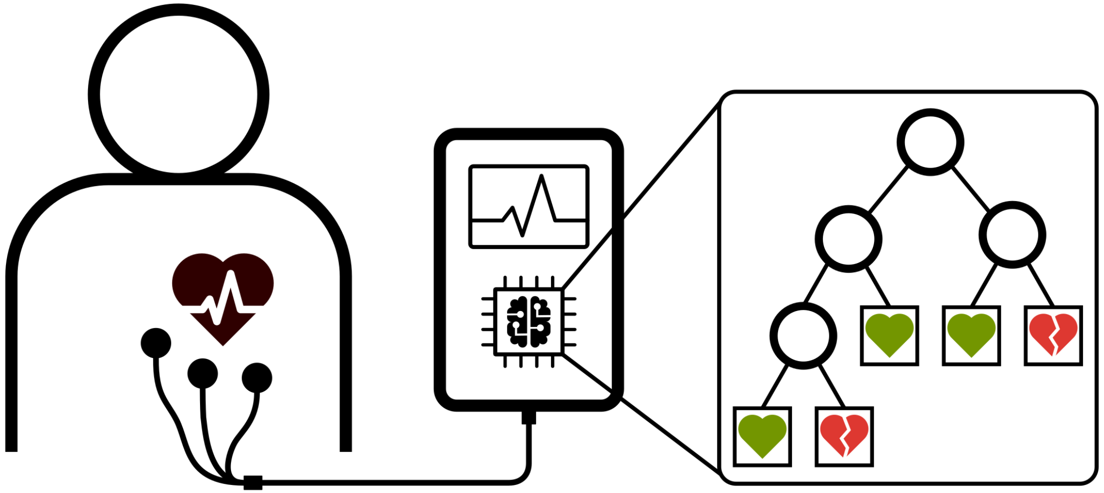
		Ground Truth	
		sick	*healthy*
Prediction	*Sick*	True Positive (TP)	False Positive (FP)
	Healthy	False Negative (FN)	True Negative (TN)

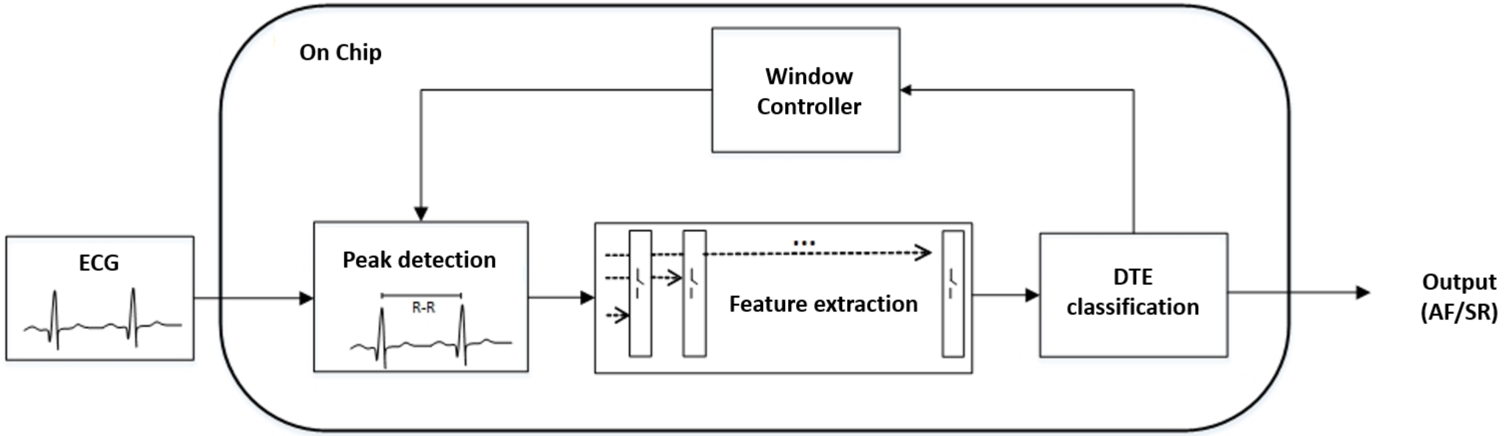
- Achieved results
  - Recall 92.7%
  - Fallout 14,7%
  - 42 nJ per 2 minute ECG-dataset

	Prediction Sinus Rhythm	Prediction Atrial Fibrillation
Sinus Rhyth	1209	209
Atrial Fibrillation	103	1315

GEnERIC

- **Concept Decision Tree**
  - Classical machine learning
    - Processing features, not raw data
    - Multi stage procedure
  - Offers potential for optimization
    - Software-Hardware implementation feedback
  - Enables utilization of problem specific expertise
  - Classification is understandable





Feature analysis

Wavelet transformation

- Achieved required accuracy
- Very costly
- Huge featureset (100~ Features)
- Big ensembles necessary (300 Trees)

Feature calculation on raw data

- Features in time and frequency domain
- Could not achieve required accuracy

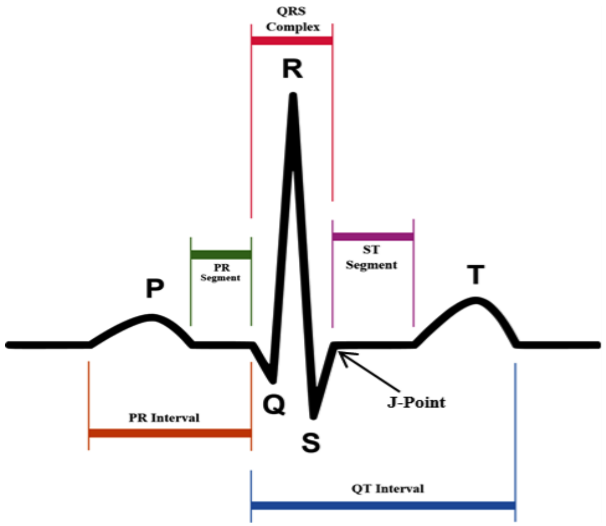
➢ Some kind of (cheap) preprocessing necessary

Feature definition

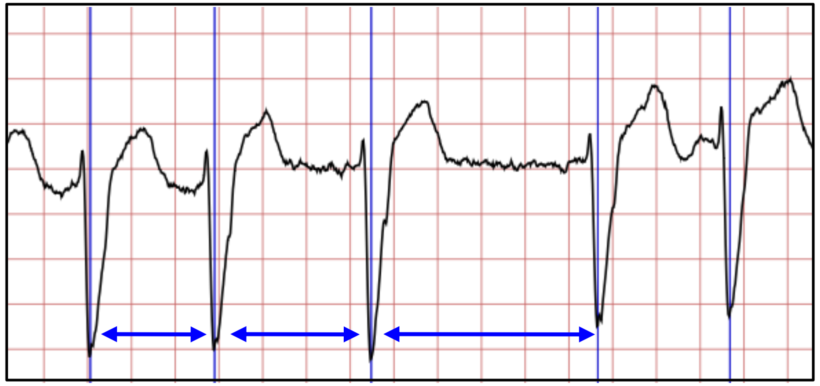
Waveform characterisitics

- Existence of P-wave
- Width of QRS-complex
- Flickering
- No promising initial results

Features defined solely based on RR-intervals



https://www.aclsmedicaltraining.com/basics-of-ecg/

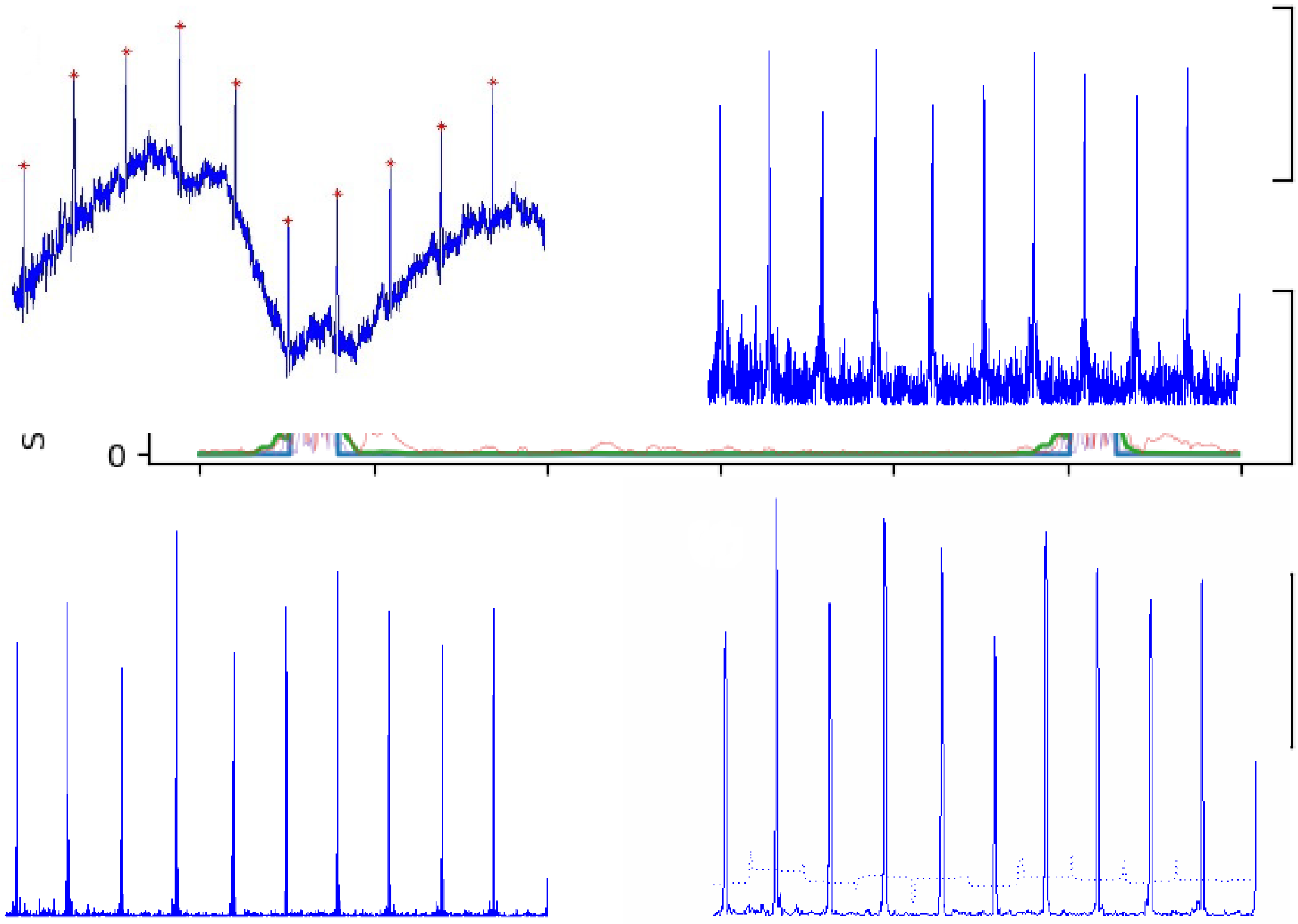


Peak Detection

Triangle Template Matching

- Highpass filter
- Template matching
- Lowpass filter
- Moving average

➢ Moving average and lowpass filtered data define search window

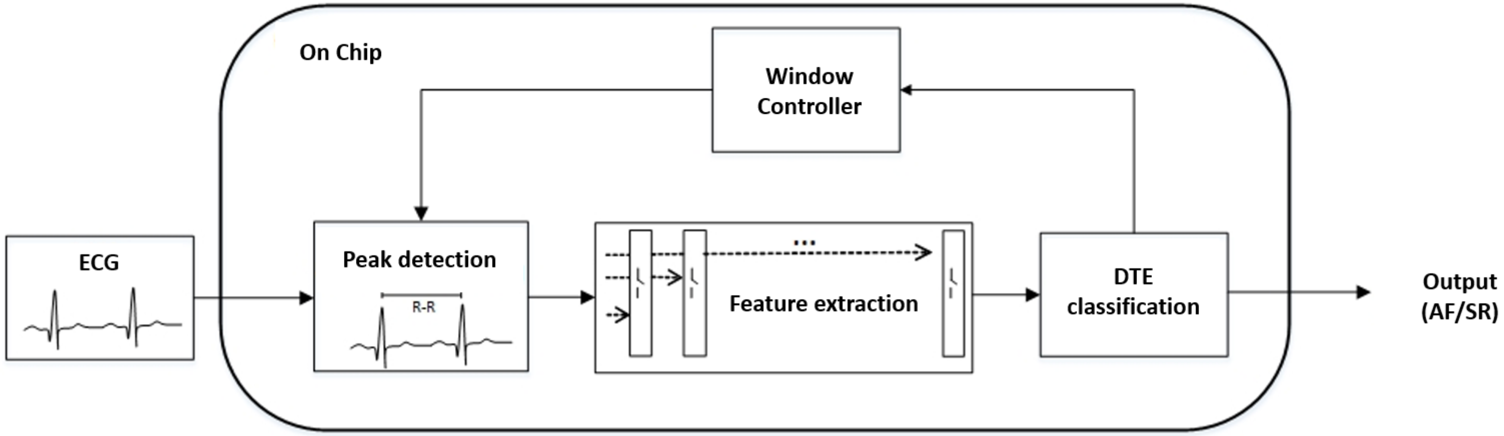


Nguyen, Tam, et al. "Low resource complexity R-peak detection based on triangle template matching and moving average filter." *Sensors* 19.18 (2019): 3997.

Optimization

Two overarching goals

- Reaching specified precision targets
- Minimizing energy cost of hardware

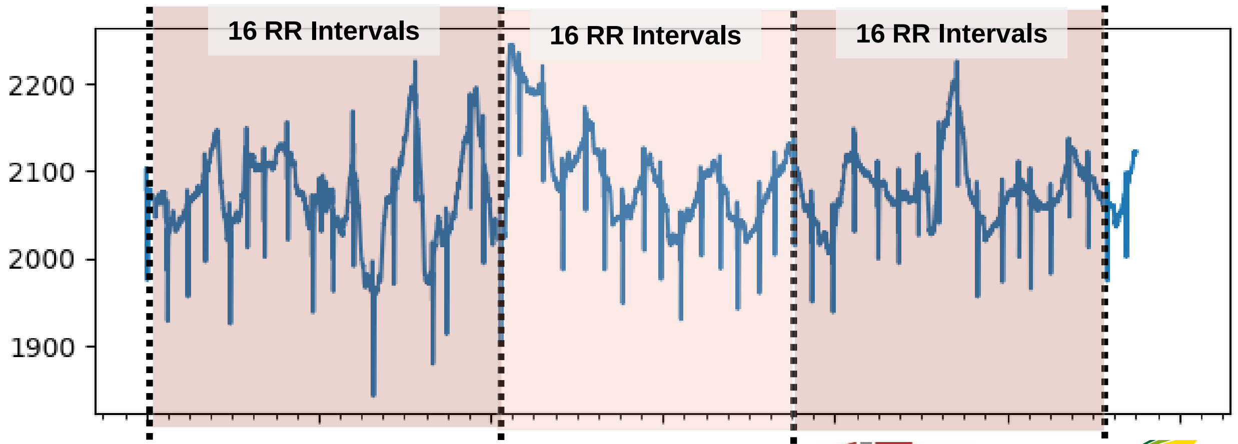


Bounded optimization problem

- Define cost function for hardware energy
  - Minimize cost function bounded by precision targets

Windowing

- Sequential classification on data windows containing 16 RR-intervals
- Enables termination on recognized fibrillation event



Optimization

Energy cost estimations

- FPGA simulation (Vivado)
  - Feature calculation
  - Decision tree ensemble

Optimization function

- Minimize energy costs
- Assign high cost to solutions not fullfilling recognition requirements

Energy Analysis						
IDX	Top Level	Modules	Sub-Modules	Dynamic Power (W) SAIF	CLK Cycles	Energy (nJ / feature)
1	fe_top_level	XXXXXXXXXXX	XXXXXXXXXXX	0,059937082	XXXXXXXXXXX	XXXXXXXXXXX
2		fe (fe)	XXXXXXXXXXX	0,030463446	XXXXXXXXXXX	XXXXXXXXXXX
3		XXXXXXXXXXX	skew_interval (fe_skew)	0,004204817	469	19,72059228
4		XXXXXXXXXXX	kurtosis_interval (fe_kurtosis)	0,004177229	469	19,59120626
5		XXXXXXXXXXX	fe_entropy (fe_entropy)	0,003984076	294	11,71318199
6		XXXXXXXXXXX	variance_interval (fe_variance)	0,002611476	177	4,62231169
7		XXXXXXXXXXX	fe_heart_rate (fe_heart_rate)	0,002534213	178	4,510899158
8		XXXXXXXXXXX	mean_interval (fe_mean)	0,002000995	177	3,541761115
9		XXXXXXXXXXX	fanofactor_interval (fe_fanofactor)	0,001934599	131	2,534324405
10		XXXXXXXXXXX	xu_3_interval (fe_xu_3)	0,001690764	30	0,507229124
11		XXXXXXXXXXX	xu_4_interval (fe_xu_4)	0,001680682	30	0,504204689
12		XXXXXXXXXXX	xu_2_interval (fe_xu_2)	0,001633506	30	0,490051776
13		XXXXXXXXXXX	xu_interval (fe_xu)	0,001492237	2	0,029844732
14		XXXXXXXXXXX	interval_shift_reg (n_shift_reg)	0,001417921	XXXXXXXXXXX	XXXXXXXXXXX
15		XXXXXXXXXXX	minmax_interval_diff (fe_minmax_diff)	0,000679171	30	0,203751441
16		XXXXXXXXXXX	std_interval (fe_std)	0,000232727	18	0,041890837
17		XXXXXXXXXXX	median_interval (fe_median)	0,000189037	30	0,056711237
18		fe_diff (fe_diff)	XXXXXXXXXXX	0,028960742	XXXXXXXXXXX	XXXXXXXXXXX
19		XXXXXXXXXXX	skew_interval_diff (fe_skew_diff)	0,004588762	469	21,52129518
20		XXXXXXXXXXX	kurtosis_interval_diff (fe_kurtosis_diff)	0,004193011	469	19,66522037
21		XXXXXXXXXXX	fe_entropy_diff (fe_entropy_diff)	0,004019144	294	11,81628197
22		XXXXXXXXXXX	variance_interval_diff (fe_variance_diff)	0,002897032	177	5,127745941
23		XXXXXXXXXXX	mean_interval_diff (fe_mean_diff)	0,001994006	177	3,529391214
24		XXXXXXXXXXX	fanofactor_interval_diff (fe_fanofactor_diff)	0,001848994	131	2,422182061
25		XXXXXXXXXXX	xu_4_interval_diff (fe_xu_4_diff)	0,001689118	30	0,506735395
26		XXXXXXXXXXX	xu_3_interval_diff (fe_xu_3_diff)	0,001643902	30	0,493170507
27		XXXXXXXXXXX	xu_2_interval_diff (fe_xu_2_diff)	0,001635007	30	0,490502093
28		XXXXXXXXXXX	xu_interval_diff (fe_xu_diff)	0,001492217	2	0,029844348
29		XXXXXXXXXXX	interval_diff_shift_reg (n_shift_reg)	0,001417915	XXXXXXXXXXX	XXXXXXXXXXX
30		XXXXXXXXXXX	interval_difference (fe_interval_diff)	0,00043447	3	0,013034098
31		XXXXXXXXXXX	fe_nn_diff_adapt (fe_adapt_nn_diff)	0,000300961	29	0,087278604
32		XXXXXXXXXXX	std_interval_diff (fe_std_diff)	0,000232726	30	0,069817717
33		XXXXXXXXXXX	median_interval_diff (fe_median_diff)	0,000148008	18	0,026641468
34		XXXXXXXXXXX	fe_nn_diff_50 (fe_nn_diff)	0,000141855	30	0,042556638
35		XXXXXXXXXXX	fe_nn_diff_25 (fe_nn_diff)	0,000141842	30	0,042552657
36		XXXXXXXXXXX	fe_nn_diff_100 (fe_nn_diff)	0,000141771	30	0,042531396
37		fe_ctrl_unit (fe_ctrl_unit)		9,77142E-05	XXXXXXXXXXX	XXXXXXXXXXX
Power Estimation Confidence Level	Medium					
Tclk	10	**ns**				
Frequency	100	**MHz**				
window length	15					
interval bitwidth	32					
feature bitwidth	32					
sum bitwidth	36					
n_bins	10					
window_length_bitwidth	3					

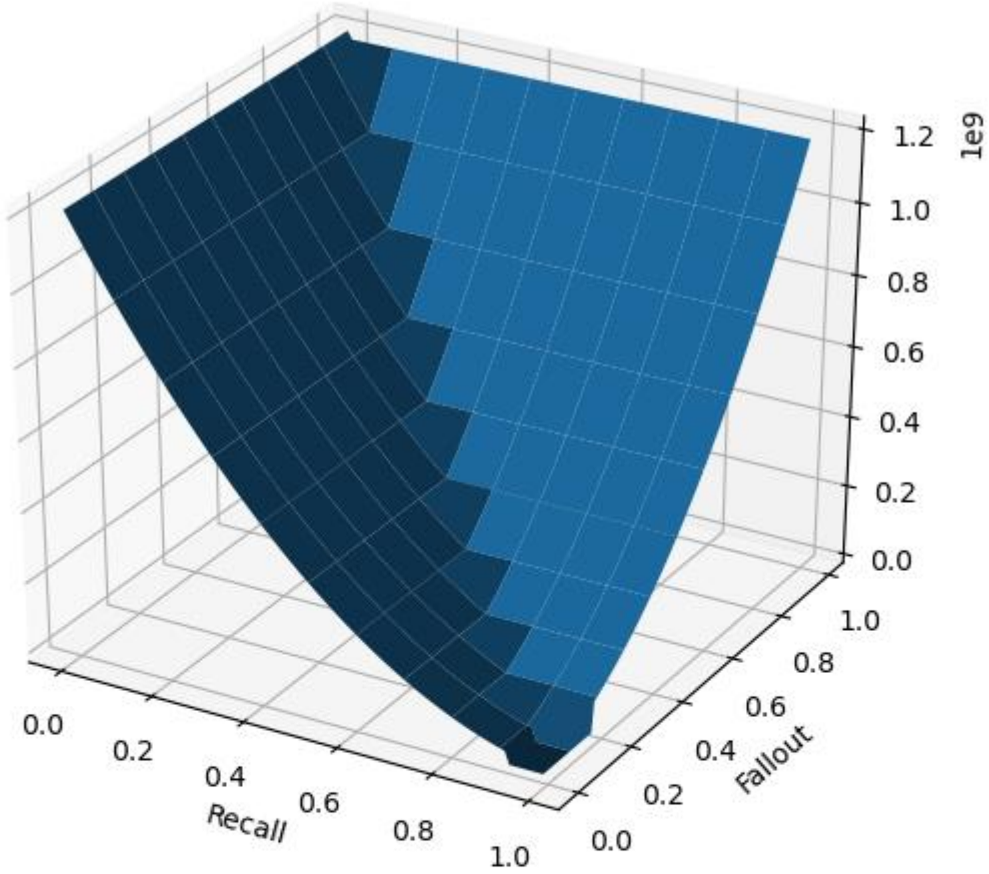
Cost function

Assign high cost to „bad“ solutions

- Recall <= 90%
- Fallout >= 20%

Assign no cost to good solutions

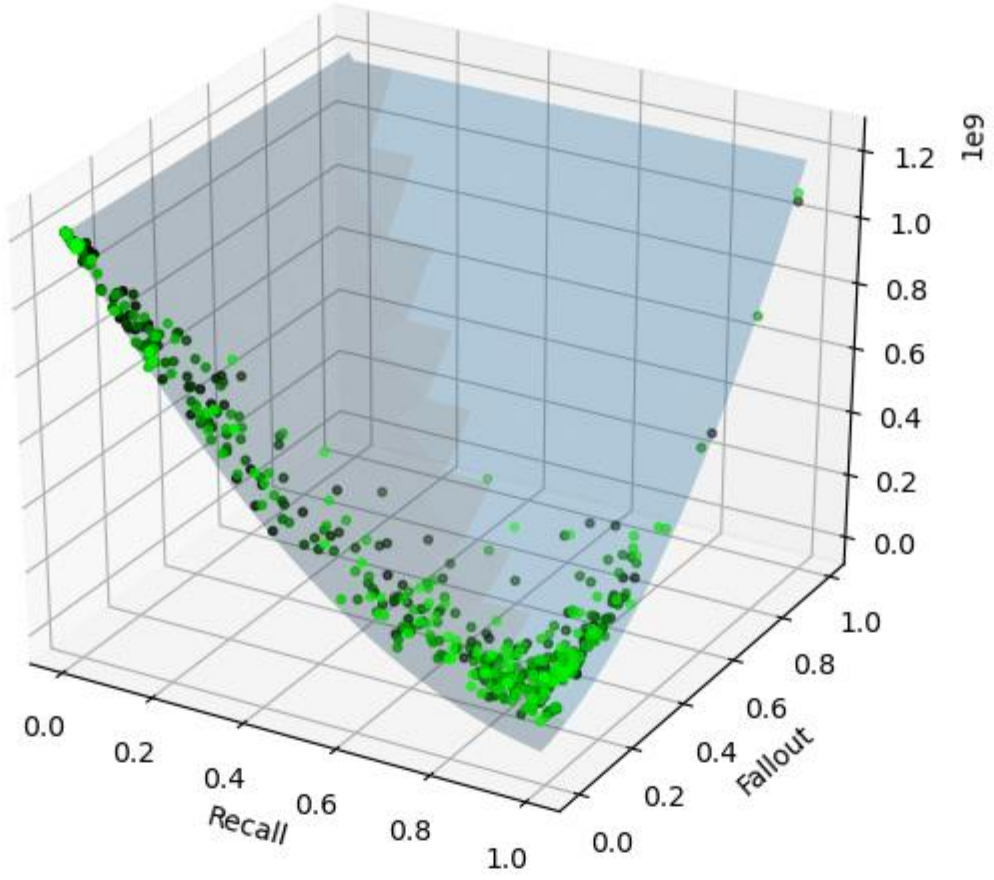
- Recall > 90%
- Fallout < 20%



Optimization

Usage of hyperopt library

- https://github.com/hyperopt/hyperopt
- Allows definition of arbitrary cost functions
- Min(Cost + Energy)



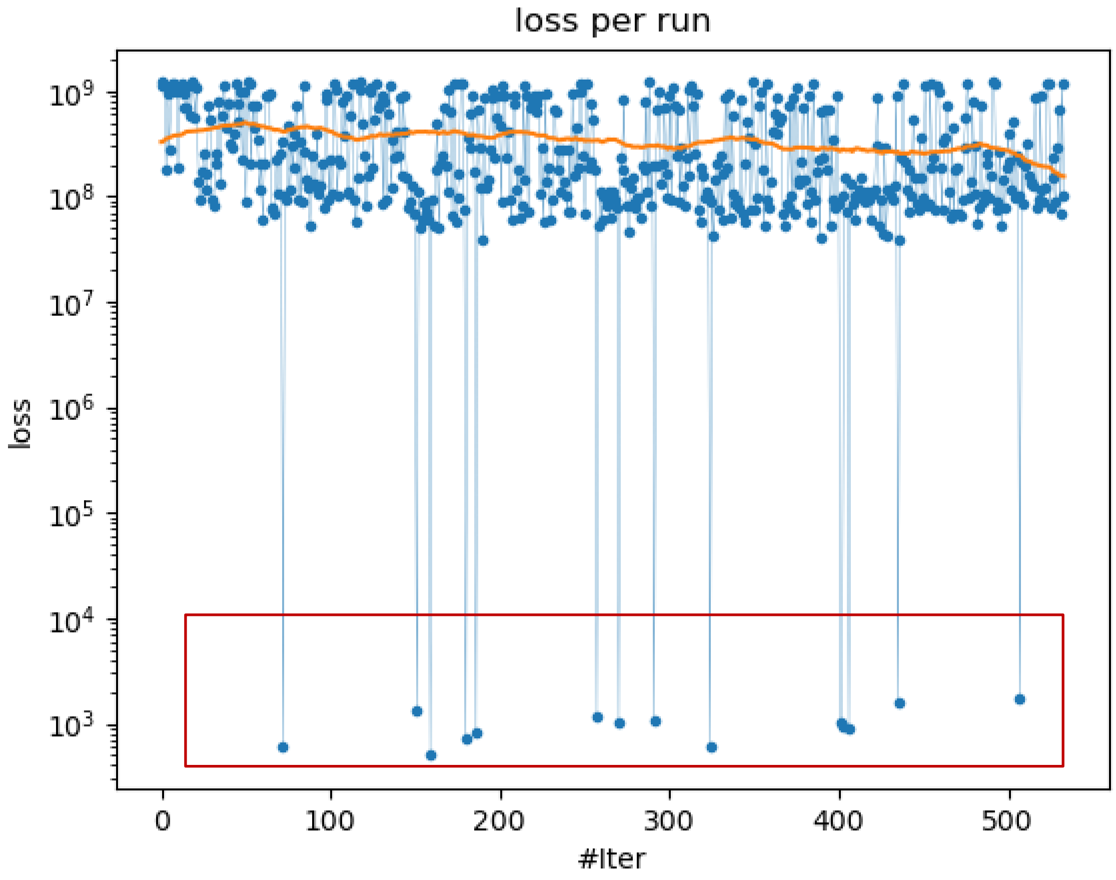
Optimization results

Sparse solutions found

- 70~ Parameters
  - Complex parameterspace
- Parameters of peak detection have huge impact
- Small amount of iterations

Used solutions as base for further improvements

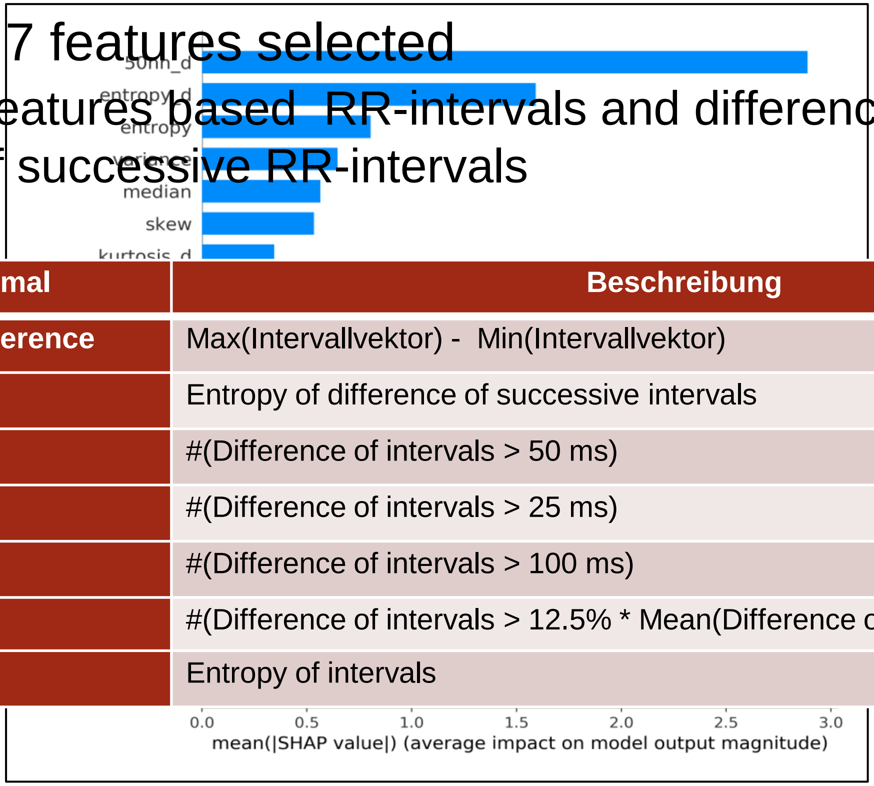
- Verification of parameter selection
  - Features
  - Decision tree ensemble
- Improvement of peak detection



Feature Selection

Derivation of several features based on RR-intervals

- Analysis of different features
- Set 7 features selected
  - ➢ Features based RR-intervals and difference of successive RR-intervals



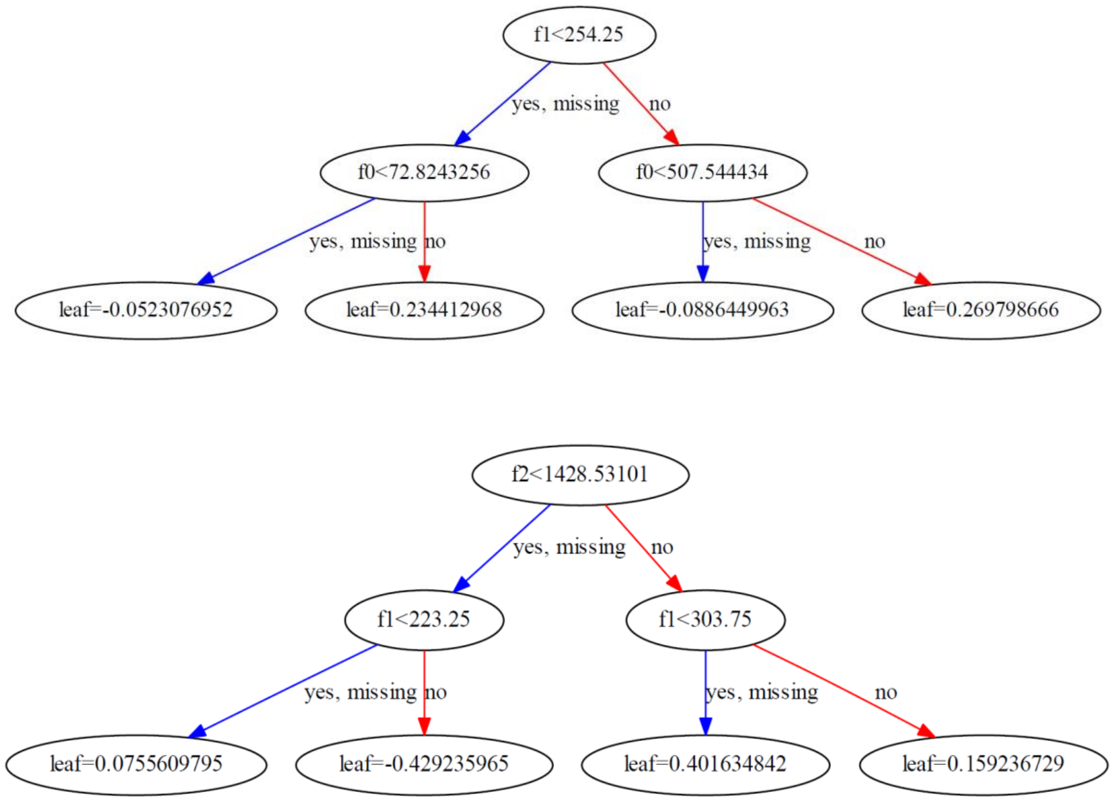
Merkmal	Beschreibung
Min-Max Difference	Max(Intervallvektor) - Min(Intervallvektor)
Entropie	Entropy of difference of successive intervals
50 nn_d	#(Difference of intervals > 50 ms)
25 nn_d	#(Difference of intervals > 25 ms)
100 nn_d	#(Difference of intervals > 100 ms)
Adapt_nn_d	#(Difference of intervals > 12.5% * Mean(Difference of Intervals))
Entropie_d	Entropy of intervals

Feature	Importance	Energy-Effiyiency	Importance	Energy [nJ / Feature]
50 NN_d	1,00	1,00	2,75	0,04
25 NN_d	0,91	1,00	2,5	0,04
100 NN_d	0,91	1,00	2,50	0,04
Adaptive NN_d	0,91	1,00	2,5	0,08
Median	0,25	1,00	0,68	0,06
Variance	0,29	0,79	0,8	4,62
MinMaxDiff	0,10	0,98	0,27	0,50
Fano Factor_d	0,09	0,89	0,25	2,42
Entropy_d	0,51	0,45	1,40	11,82
Fano Factor	0,07	0,88	0,2	2,53
Heartrate	0,02	0,79	0,05	4,51
Entropy	0,33	0,46	0,9	11,71
Variance_d	0,02	0,76	0,05	5,13
Skew	0,24	0,08	0,65	19,72
Kurtosis_d	0,11	0,09	0,30	19,67
Kurtosis	0,05	0,09	0,15	19,59
Skew_d	0,02	0,00	0,05	21,52

Classification

Decision tree ensemble

- Multiple decision trees
- Sum of leafs defines classification result
- Gradient boosting algorithm



Parameters

- 20 Trees
- Depth of 3
- 10 Bins

Voting

Each tree has a binary vote (0,1) → [1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1]

- Each vote gets multiplied by some tree weight → [10, 8, 7, 7, 5, 6, 6, 4, 5, 4, 4, 3, 4, 3, 3, 2, 3, 3, 5, 2]
- Sum of all votes
- If votes surpass certain threshhold
  - ➢ Atrial fibrillation
  - ➢ Else: Sinus Rhythm

Average of absolute leaf values determines tree weight

➢ Large leaf values have huge impact on classification

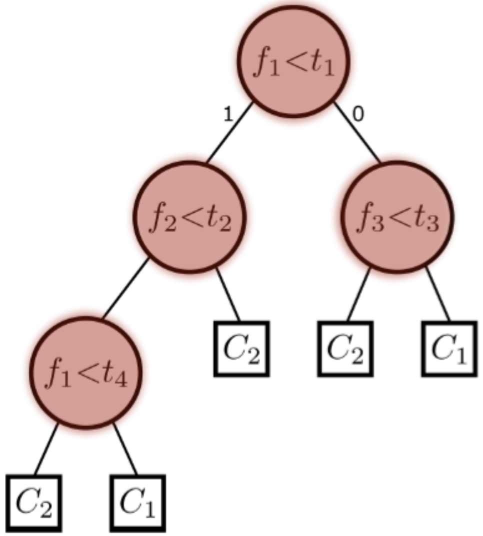
Classification in Hardware

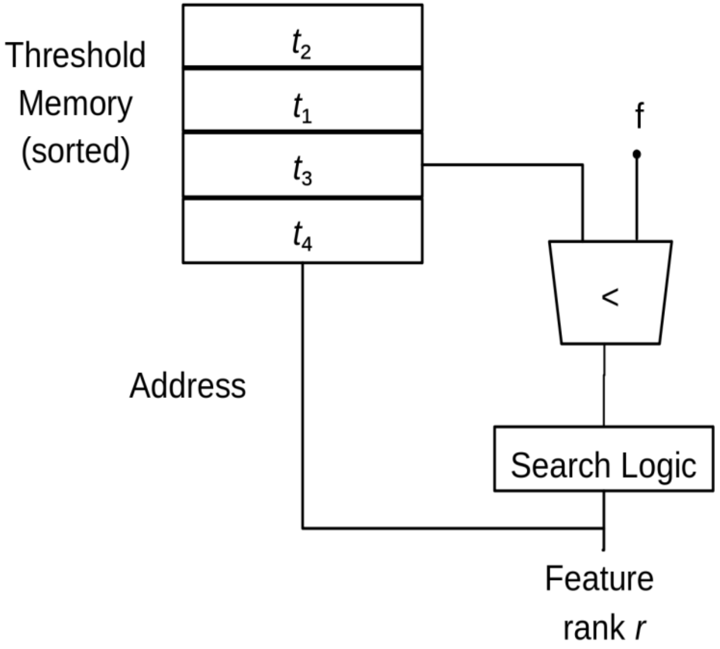
Evaluation of trees in HW

- Variable structure
- Complex interconnect
- Redundant comparisons

Rank instead of feature value

- Fewer comparisons
- Lower data width





Classification in Hardware

Patent pending!

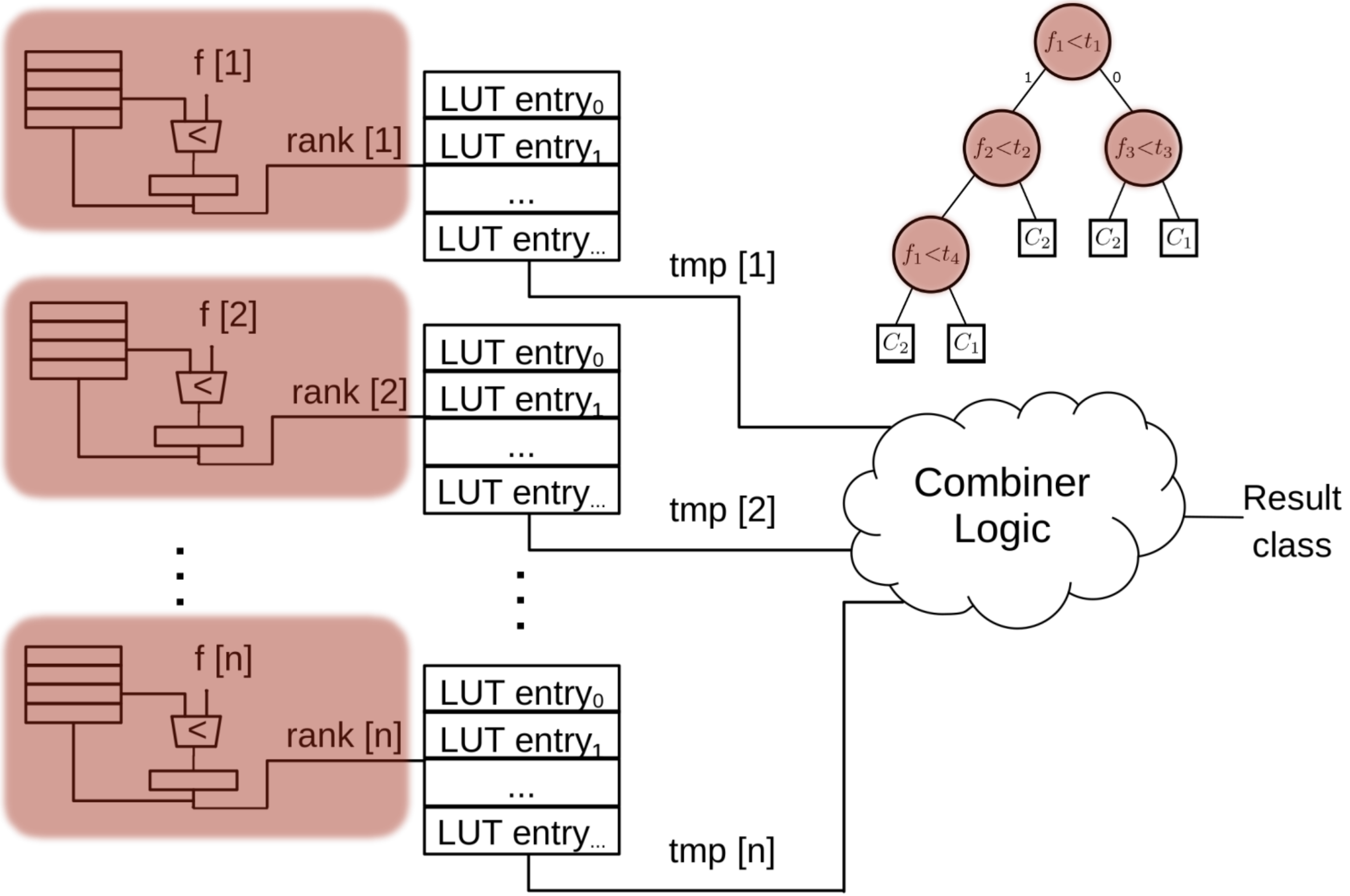
Ensemble reordered based on features instead of trees

Implementation

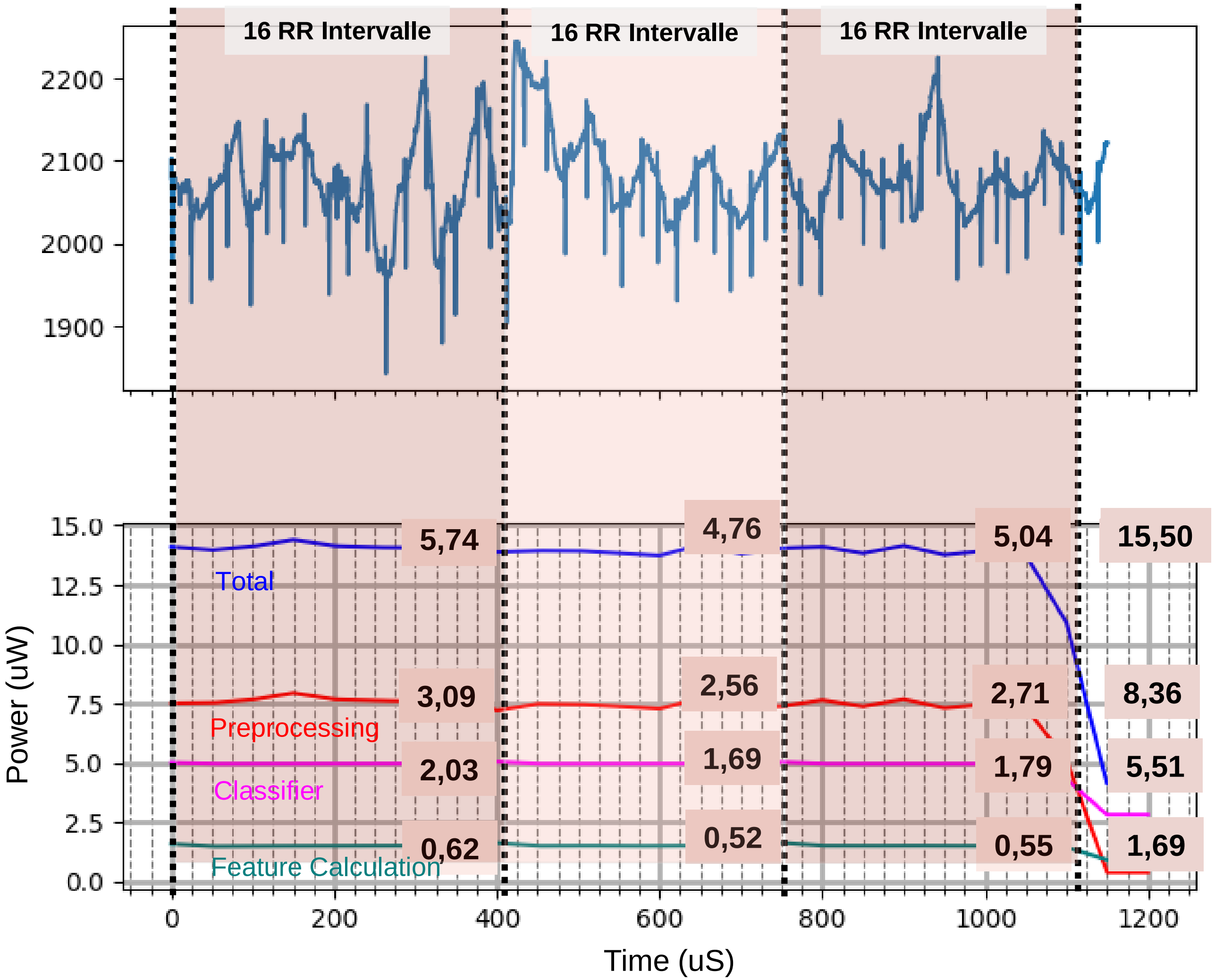
- CAM-LUT architecture
- Pre-calculation of partial results
- Compression rates of up to 400 %
- Fetch from memory
- Small logic
- Completely reconfigurable

Combiner

- Weighting of trees
- Threshold for termination criterion

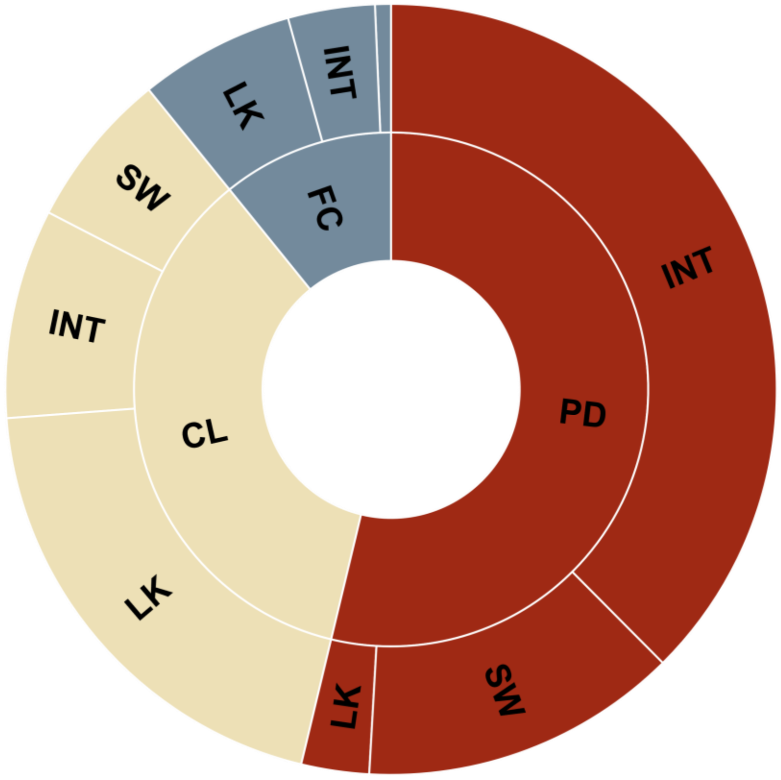


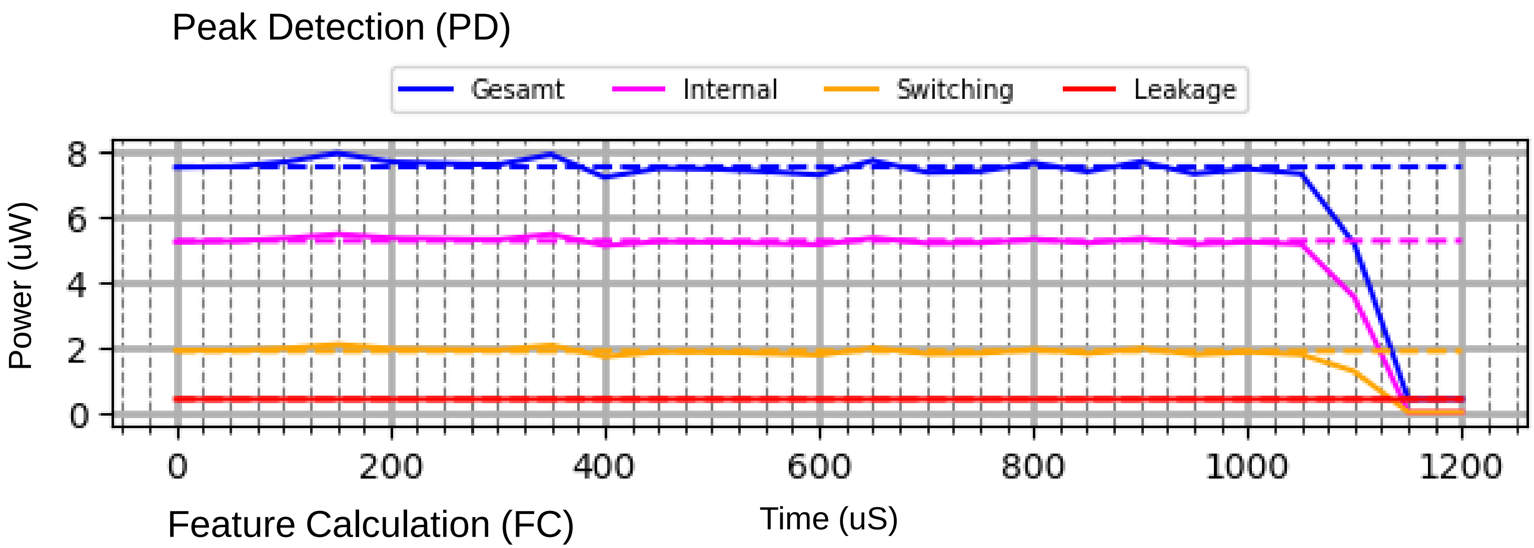
Simulation results

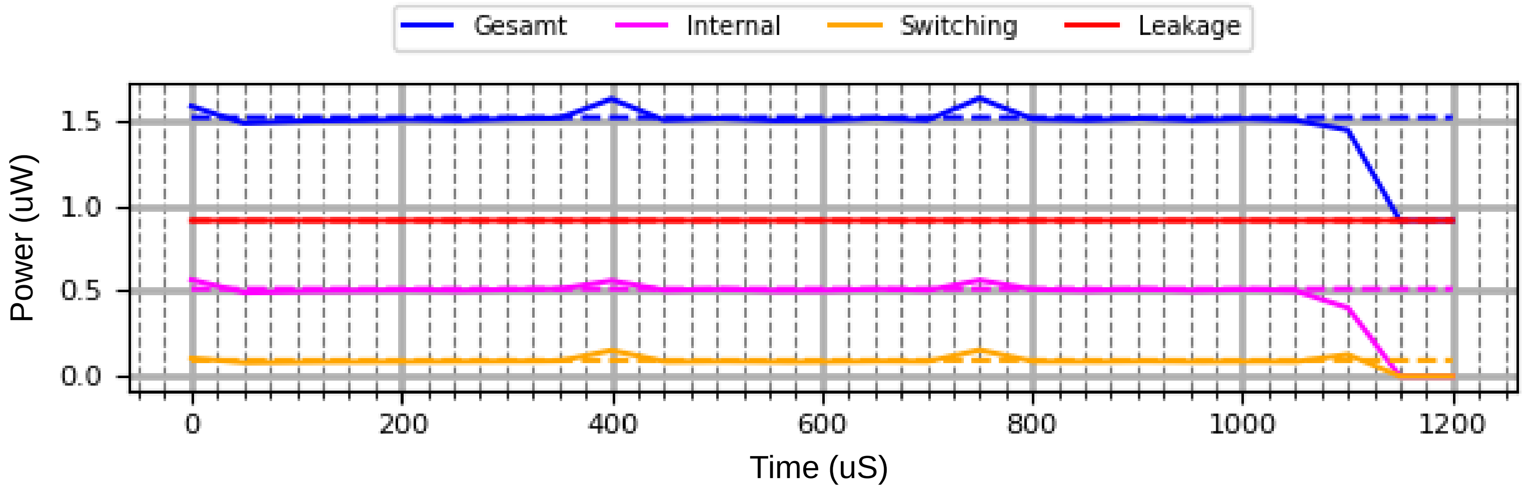


	Testset w/o termination	Testset With termination
Number of files	2836	
Mean #windows per file	7,8	6,3
Energy per file [nJ]	42,0	32,5
Recall	92,5	92,7
Fallout	14,4	14,7

Energy Breakdown







Conclusion

Classical ML approach

- Efficiency instead of number crunching
- Understandable classification

HW/SW Co-Optimization

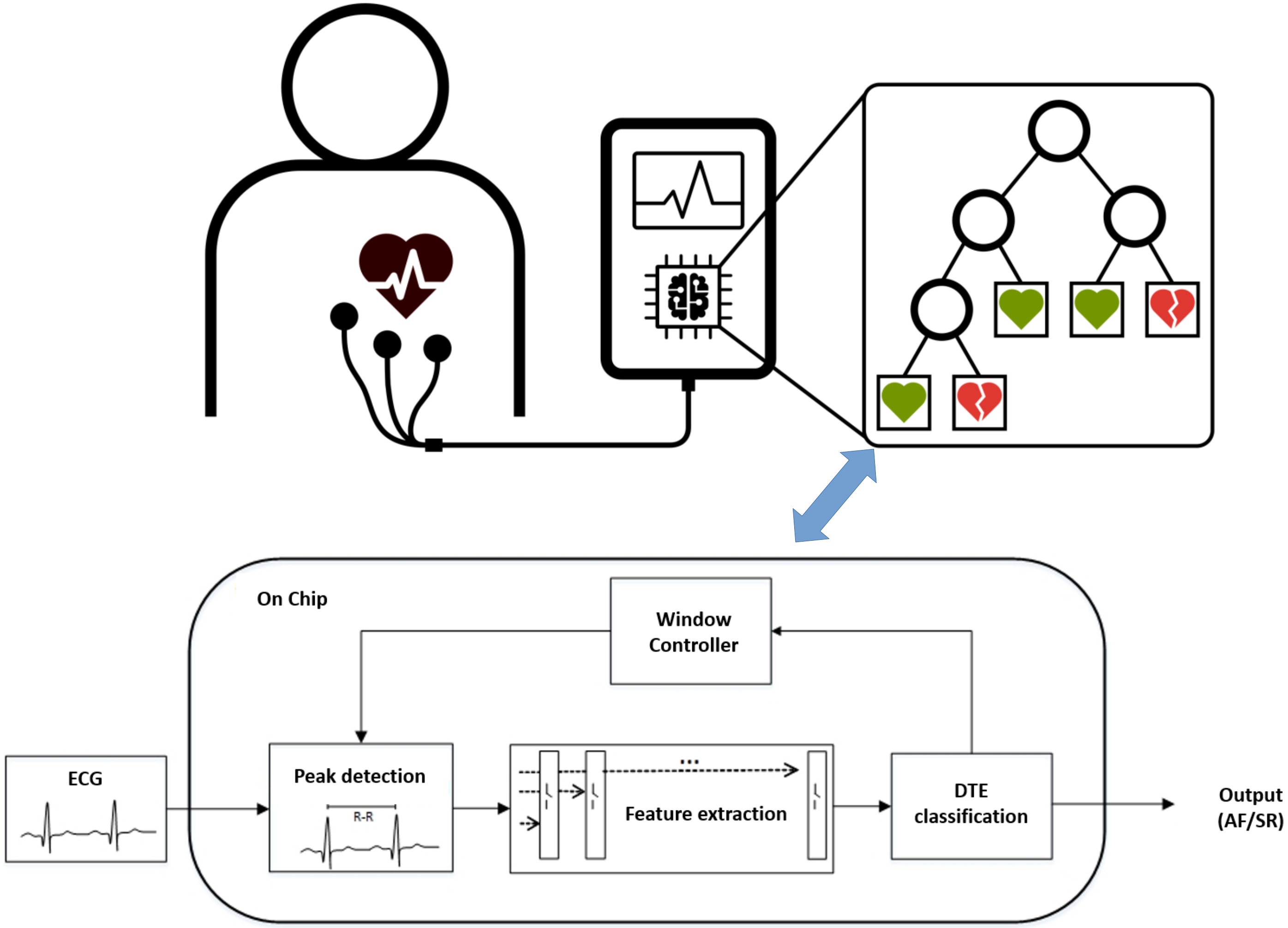
- SW model using HW cost estimation
- HW optimized for SW model

Innovative classifier

- Rank calculation
- CAM-LUT Architecture

Low energy solution

- max. 42,0nJ / 2 minute dataset
- typ. 32,5nJ / dataset



Hahn Schickard

Visions to Products

Thanks for your attention!

Copyright Notice

www.tinyml.org

Copyright Notice

www.tinyML.org