ML: Research and Applications

Insys Project

Dr. Mariela De Lucas Alvarez

DFKI - Robotics Innovation Center
Interaction and Learning

Director: Prof. Dr. Frank Kirchner

https://robotik.dfki-bremen.de/
robotik@dfki.de

January 23, 2024

Universität Bremen

Overview

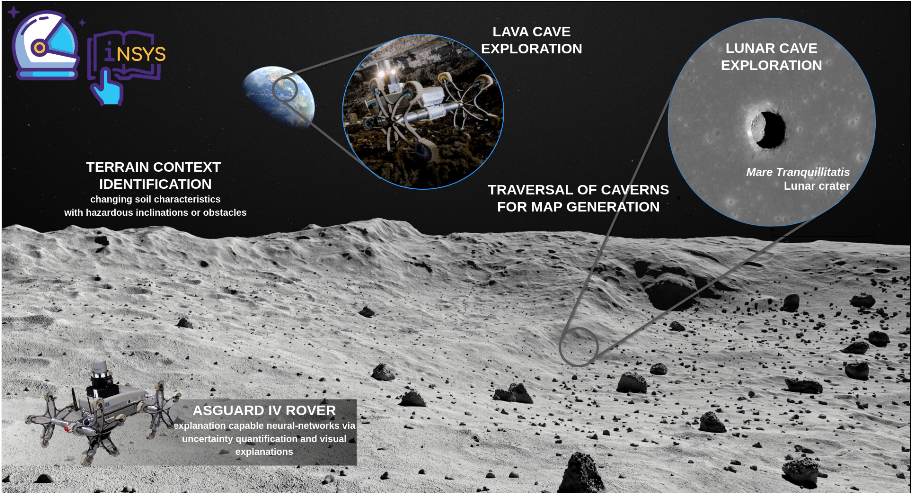
Interpretable Monitoring Systems

The goal of INSYS is to develop Explainable Artificial Intelligence (XAI) methods for multi modal robotic systems in the context of space applications.

XAI is the research field concerned with ([Langer2021, Gilpin2019]):

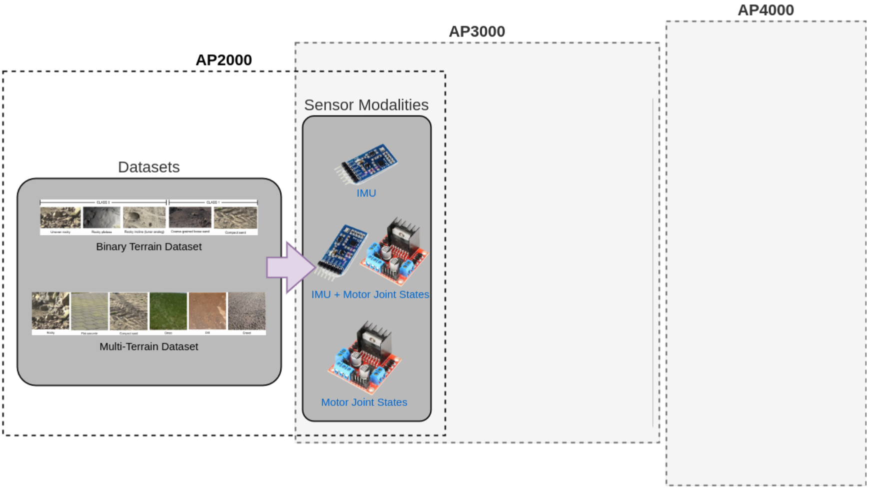
- Development of approaches that make AI understandable to humans.
- Providing insights into behaviour and processes of complex systems.
- Allowing part of the internal system to be more transparent.

Insys Objectives



Universität Bremen

Multimodal Data Aquisition and Processing



Datasets

- DFKI-RIC Lunar Cave Analog
  - Recordings of mobility hazards
  - Rock terrain with 30° slopes
- Entern Trials: Canary Islands
  - Lava tubes - Hard rocky terrain
- DFKI-RIC Test Field
  - Compact and loose sand terrain, Concrete and Grass
  - Mostly flat with some slopes

Classification

- Binary Terrain


Uneven rocky


Rocky plateau


Rocky incline (lunar analog)


Coarse-grained loose sand


Compact sand

- Multi-Category Terrain (6 classes)


Rocky

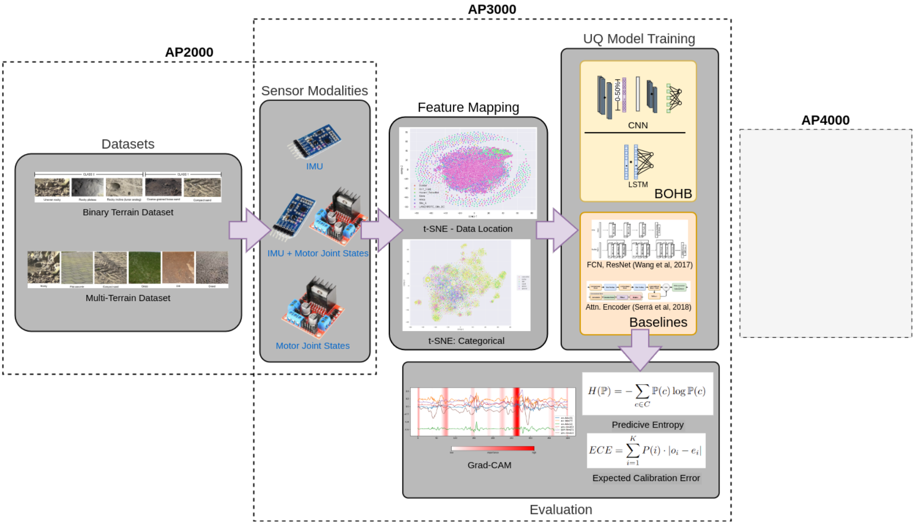

Flat concrete


Compact sand

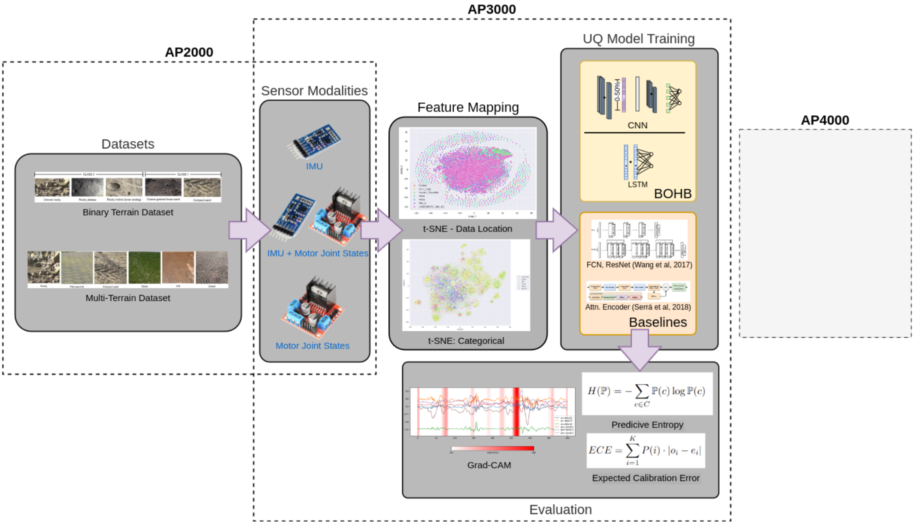

Grass


Dirt

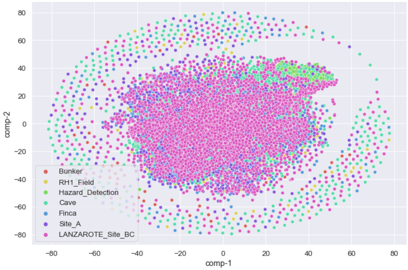

Gravel


AP3000
AP2000
AP4000
UQ Model Training
Sensor Modalities
Feature Mapping
Datasets
IMU
CNN
LSTM
BOHB
Binary Terrain Dataset
IMU + Motor Joint States
t-SNE - Data Location
FCN, ResNet (Wang et al, 2017)
Multi-Terrain Dataset
Attn. Encoder (Serrá et al, 2018)
Baselines
Motor Joint States
t-SNE: Categorical
$H(\mathbb{P}) = -\sum_{c \in C} \mathbb{P}(c) \log \mathbb{P}(c)$
Predicive Entropy
$ECE = \sum_{i=1}^{K} P(i) \cdot |o_i - e_i|$
Grad-CAM
Expected Calibration Error
Evaluation

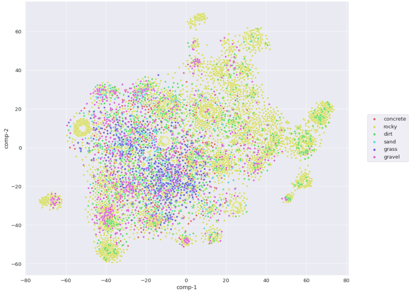
Architecture Design and Tuning



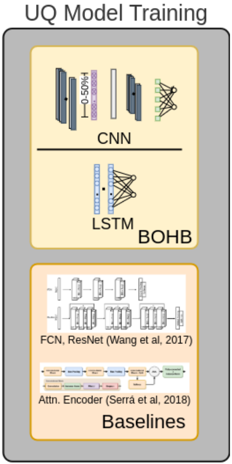
Feature Mapping: t-SNE



Location categories

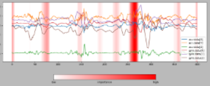


Terrain categories



- Various combinations of CNNs and Optimized tuning with BOHB (Bayesian Optimization and Hyperband) method [Falkner2018]
- Also include SOTA benchmark models: Residual Network (ResNet) [Wang 2017], Fully Convolutional Network (FCN) [Wang 2017], Encoder [Fawaz 2019]
- 3 Uncertainty Quantification methods were applied on those models: Monte Carlo Dropout [Gal 2016], Monte Carlo DropConnect [Mobiny 2021], Flipout [Wen 2018]
- In total, 480 BOHB models + 48 benchmark models with their variants = 528 models

Evaluation



Grad-CAM

$$H(\mathbb{P}) = -\sum_{c \in C} \mathbb{P}(c) \log \mathbb{P}(c)$$

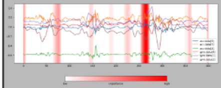
Predicive Entropy

$$ECE = \sum_{i=1}^{K} P(i) \cdot |o_i - e_i|$$

Expected Calibration Error

- Grad-CAM, a back-propagation-based XAI method, locates contributing regions in the raw data for predicting specific labels.
- Predictive Entropy to measure the uncertainty associated with a models predictions.
- Expected Calibration Error (ECE) assesses the calibration of the probabilistic model. Helps determine if the predictions align with the actual outcomes.
- Metrics and Grad-CAM used in conjunction to better understand the outputs of our classifiers with high confidence.

Evaluation



Grad-CAM

$$H(\mathbb{P}) = -\sum_{c \in C} \mathbb{P}(c) \log \mathbb{P}(c)$$

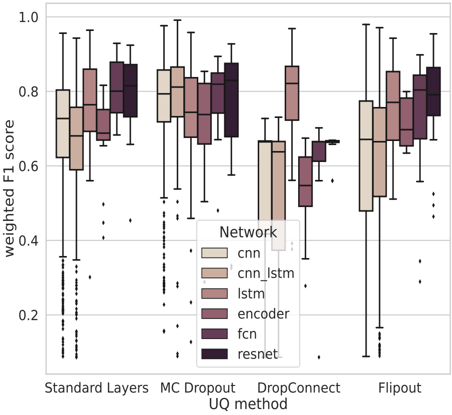
Predicive Entropy

$$ECE = \sum_{i=1}^{K} P(i) \cdot |o_i - e_i|$$

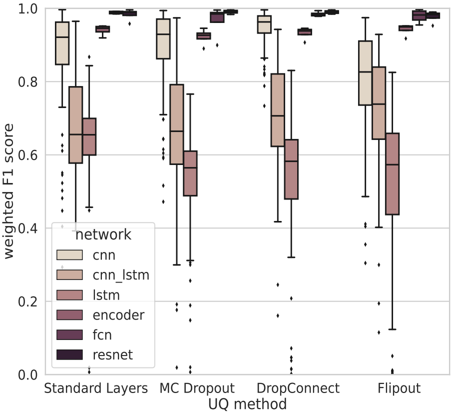
Expected Calibration Error

- Grad-CAM, a back-propagation-based XAI method, locates contributing regions in the raw data for predicting specific labels.
- Predictive Entropy to measure the uncertainty associated with a models predictions.
- Expected Calibration Error (ECE) assesses the calibration of the probabilistic model. Helps determine if the predictions align with the actual outcomes.
- Metrics and Grad-CAM used in conjunction to better understand the outputs of our classifiers with high confidence.

Weighted F1-score for all trained models (AP3200)

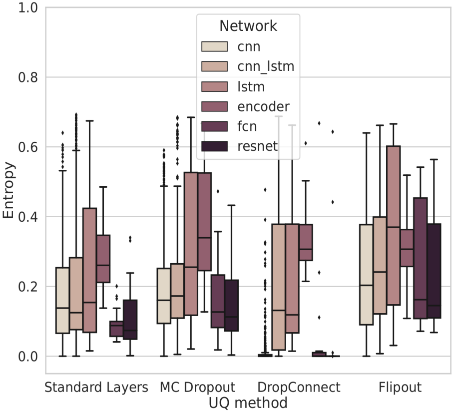


Binary terrain classification (6603 models)

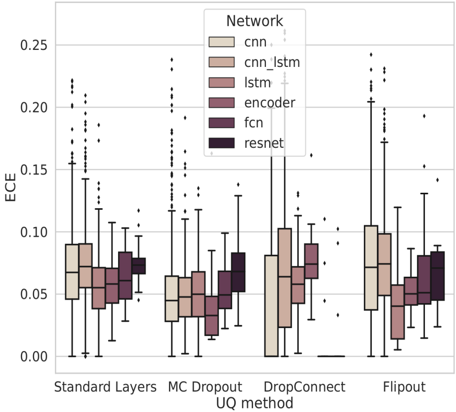


Multi-terrain classification (1268 models)

Detection of Inconsistencies - Binary Terrain (AP3400)

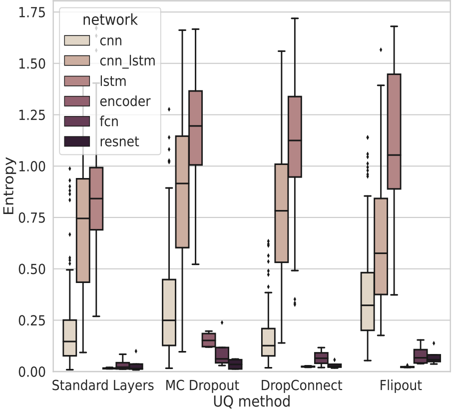


Binary Terrain Predictive Entropy

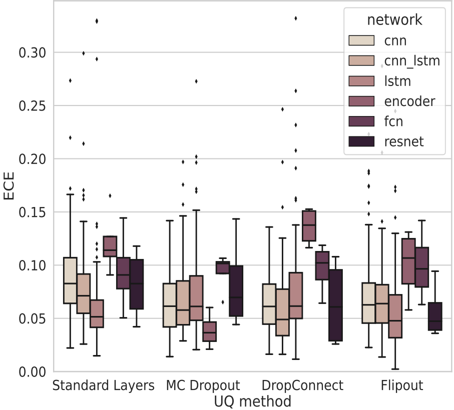


Binary Terrain ECE

Detection of Inconsistencies - Multi-Class Terrain (AP3400)



Multi-class Predictive Entropy

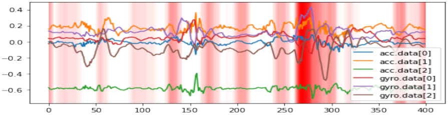


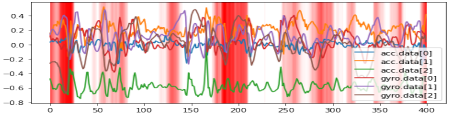
Multi-class Terrain ECE

Evaluation

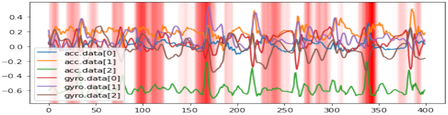
Grad-CAM Visualization [Selvaraju et al 2017]

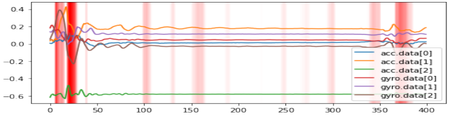
- Rock:



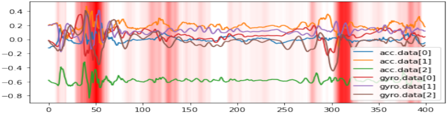


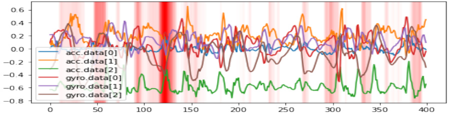
- Sand:





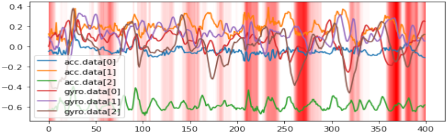
- Concrete:

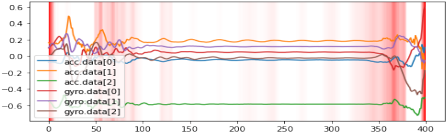




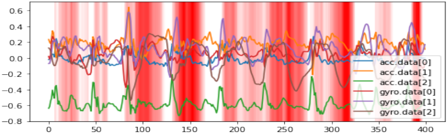
Grad-CAM Visualization

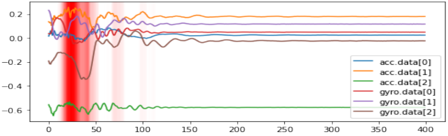
- Dirt:



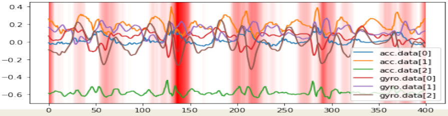


- Gravel:





- Grass:



Advantages of using XAI and UQ methods in ML applications:

Interpretability and Trust:

- **Grad-CAM**: Visualize and interpret model decisions by highlighting influential regions in input data.
- **Uncertainty Methods**: Provide insights into prediction reliability, promoting trust and cautious reliance on model outputs.

Robustness and Bias Detection:

- **Grad-CAM**: Reveals model biases through visualizing influential features, aiding bias detection.
- **Uncertainty Methods**: Assess model robustness by capturing prediction variations and identifying areas of uncertainty.

Thank you!

mariela.de_lucas_alvarez@dfki.de

References I

Langer, M. et al (2021)

What Do We Want From Explainable Artificial Intelligence? A Stakeholder Perspective on XAI and a Conceptual Model Guiding Interdisciplinary XAI Research.

https://doi.org/10.1016/j.artint.2021.103473

Gilpin, L. et al (2019)

Explaining Explanations: An Overview of Interpretability of Machine Learning

https://arxiv.org/pdf/1806.00069.pdf

Ismail, A. A., Gunady, M., Bravo, H. C., & Feizi, S. (2020).

Benchmarking Deep Learning Interpretability in Time Series Predictions.

http://arxiv.org/abs/2010.13924

O'Sullivan, Conor (2020)

Interpretable vs Explainable Machine Learning

Towards Data Science

https://towardsdatascience.com/interperable-vs-explainable-machine-learning-1fa525e12f48

References II

Saveriano, M. et al (2019)
Symbolic Task Compression in Structured Task Learning
https://doi.org/10.1109/IRC.2019.00033

Montavon, G. et al (2017)
Methods for Interpreting and Understanding Deep Neural Networks
https://arxiv.org/abs/1706.07979v1

Falkner, S. et al (2018)
BOHB: Robust and efficient hyperparameter optimization at scale
http://proceedings.mlr.press/v80/falkner18a/falkner18a.pdf

Wang, Z. et al (2017)
Timeseries classification from scratch with deep neural networks: A strong baseline
http://pmo78bb17.pic35.websiteonline.cn/upload/1611.pdf

Fawaz, H.I. et al (2019)
Deep learning for time series classification: a review
https://arxiv.org/pdf/1809.04356.pdf

References III

Zhou, B. et al (2016)
Learning Deep Features for Discriminative Localization
https://dspace.mit.edu/handle/1721.1/112986

Van der Maaten, L. & Hinton,G. (2008)
Visualizing Data using t-SNE
https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf?fbclid

Fawaz et al (2019)
ResNet model
https://github.com/hfawaz/dl-4-tsc/blob/master/classifiers/resnet.py

Fawaz et al (2019)
FCN model
https://github.com/hfawaz/dl-4-tsc/blob/master/classifiers/fcn.py

De Lucas Alvarez (2021)
Compact Neural Networks for Autonomous Robots
Dissertation for Doctoral Degree May 2021

References IV

Abdar et al (2021)
A review of uncertainty quantification in deep learning: Techniques, applications and challenges.
https://doi.org/10.1016/j.inffus.2021.05.008

Gal, Y. et al (2016)
Dropout as a bayesian approximation: Representing model uncertainty in deep learning
http://proceedings.mlr.press/v48/gal16.pdf

Mobiny, A. et al (2021)
Dropconnect is effective in modeling uncertainty of bayesian deep networks
https://www.nature.com/articles/s41598-021-84854-x

Wen W. et al (2018)
Flipout: Efficient pseudo-independent weight perturbations on mini-batches
https://arxiv.org/pdf/1803.04386.pdf

References V

Selvaraju, Ramprasaath R., et al (2017)

Grad-cam: Visual explanations from deep networks via gradient-based localization

Guo, C et al (2017)

On calibration of modern neural networks

Naeini et al (2017)

Obtaining well calibrated probabilities using bayesian binning