UNCERTAINTY WRAPPERS FOR DATA-DRIVEN MODELS

INCREASE THE TRANSPARENCY OF AI/ML-BASED MODELS THROUGH ENRICHMENT WITH DEPENDABLE SITUATION-AWARE UNCERTAINTY ESTIMATES

Turku, Finland, 10 September 2019

WAISE 2019: Second International Workshop on Artificial Intelligence Safety Engineering

Michael Kläs

Lena Sembach

Fraunhofer IESE

MInD RheinlandPfalz

CrESt

Our Focus: Data-Driven Components

Intended Functionality

f(x) = x*1.19

Specification

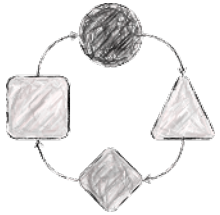
(Engineered) Component

Code

Intended Functionality

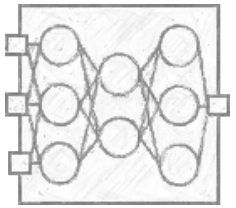
Recognize Stop Signs

Labeled Examples



Learning Approach

Data-Driven Component



Data-Driven Model

CNN

Does the component provide the intended outcome for each input ?

MInD

CrESt

Fraunhofer IESE

Information

Outcome

No stop sign

Uncertainty

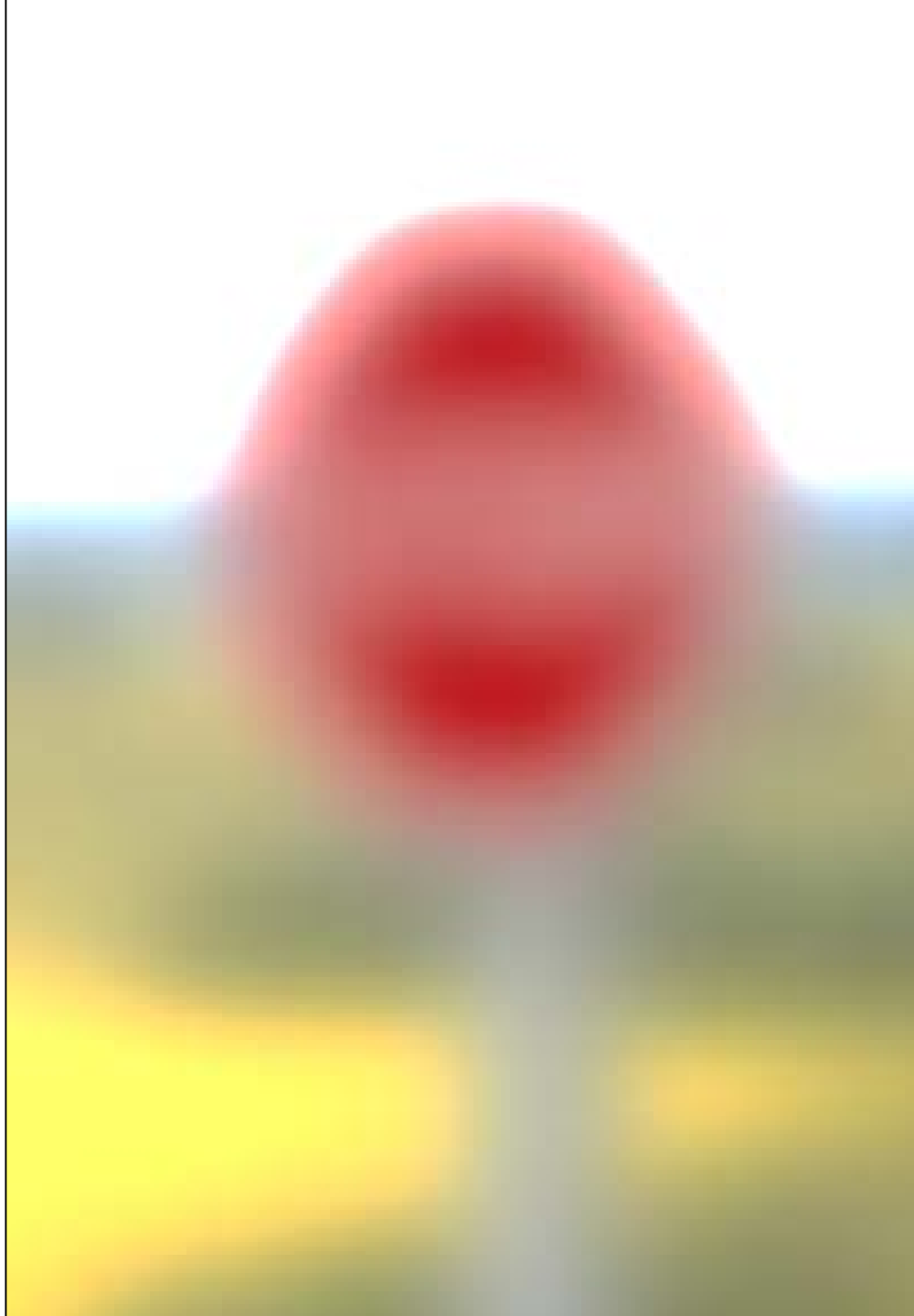
Error probability

0.5% | 40%

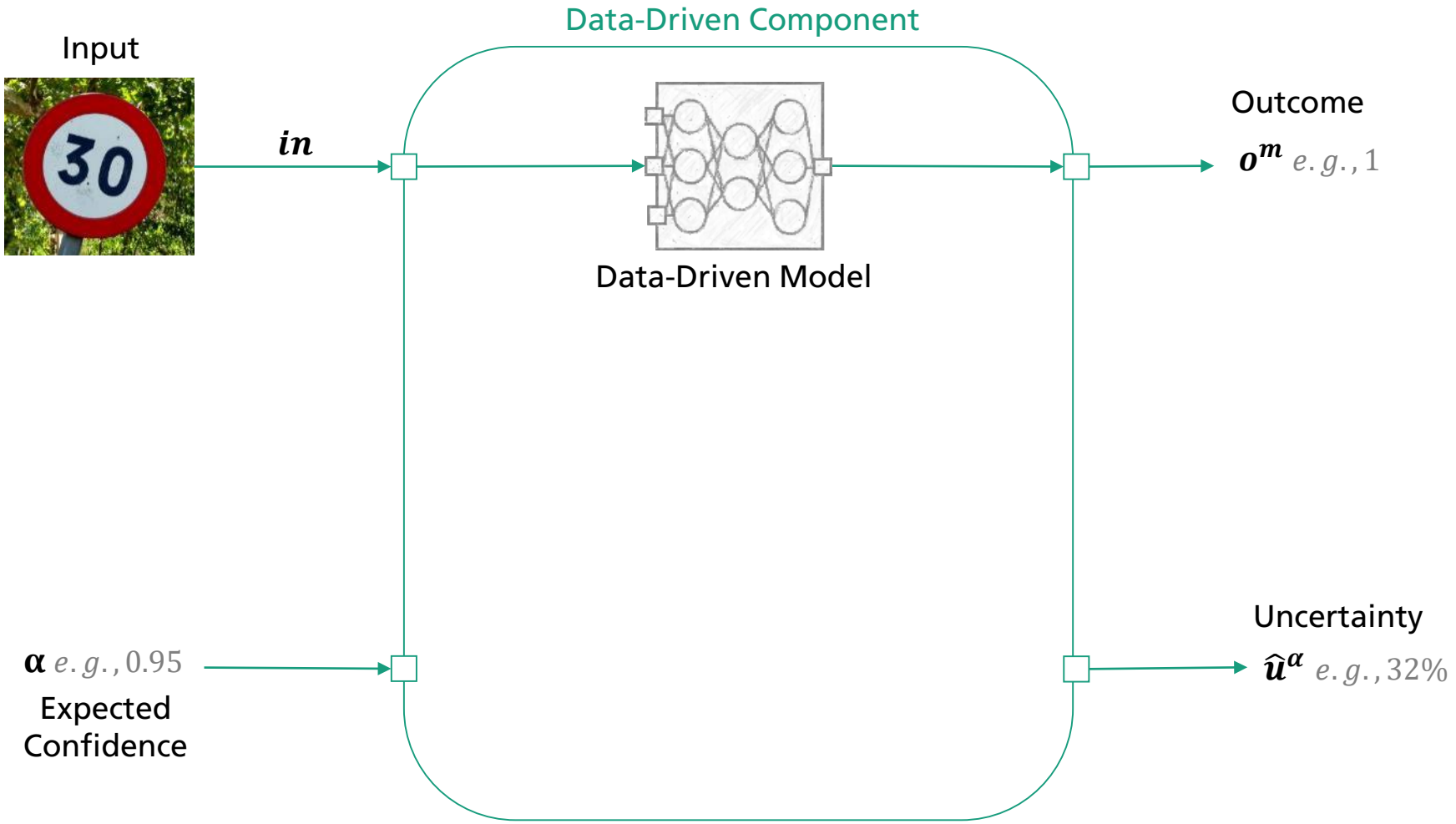
Dependability

Confidence level

95%



Uncertainty Wrapper: An Architectural Perspective

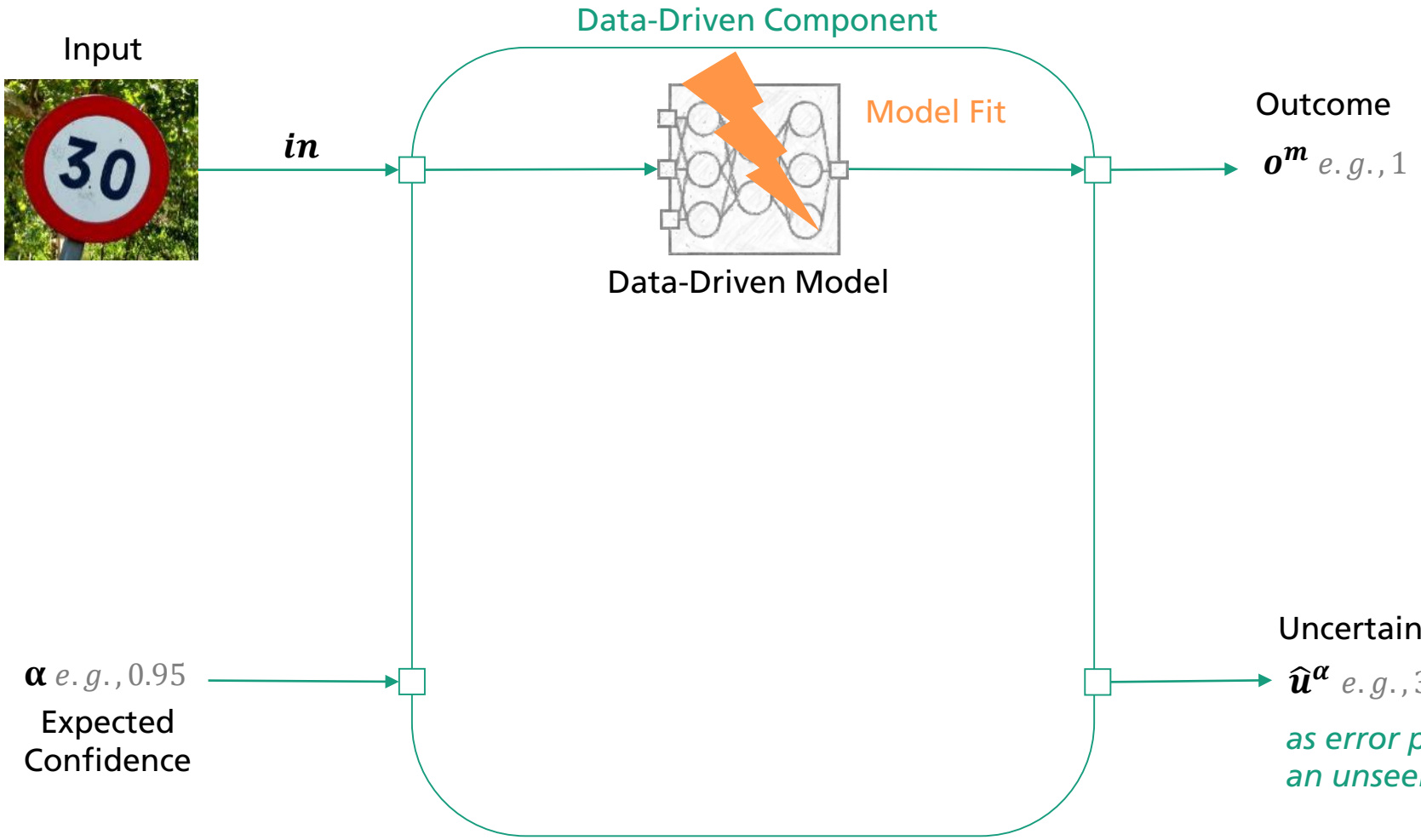


MInD RheinlandPfalz

CrESt

Fraunhofer IESE

Key Sources of Uncertainty: Model Fit

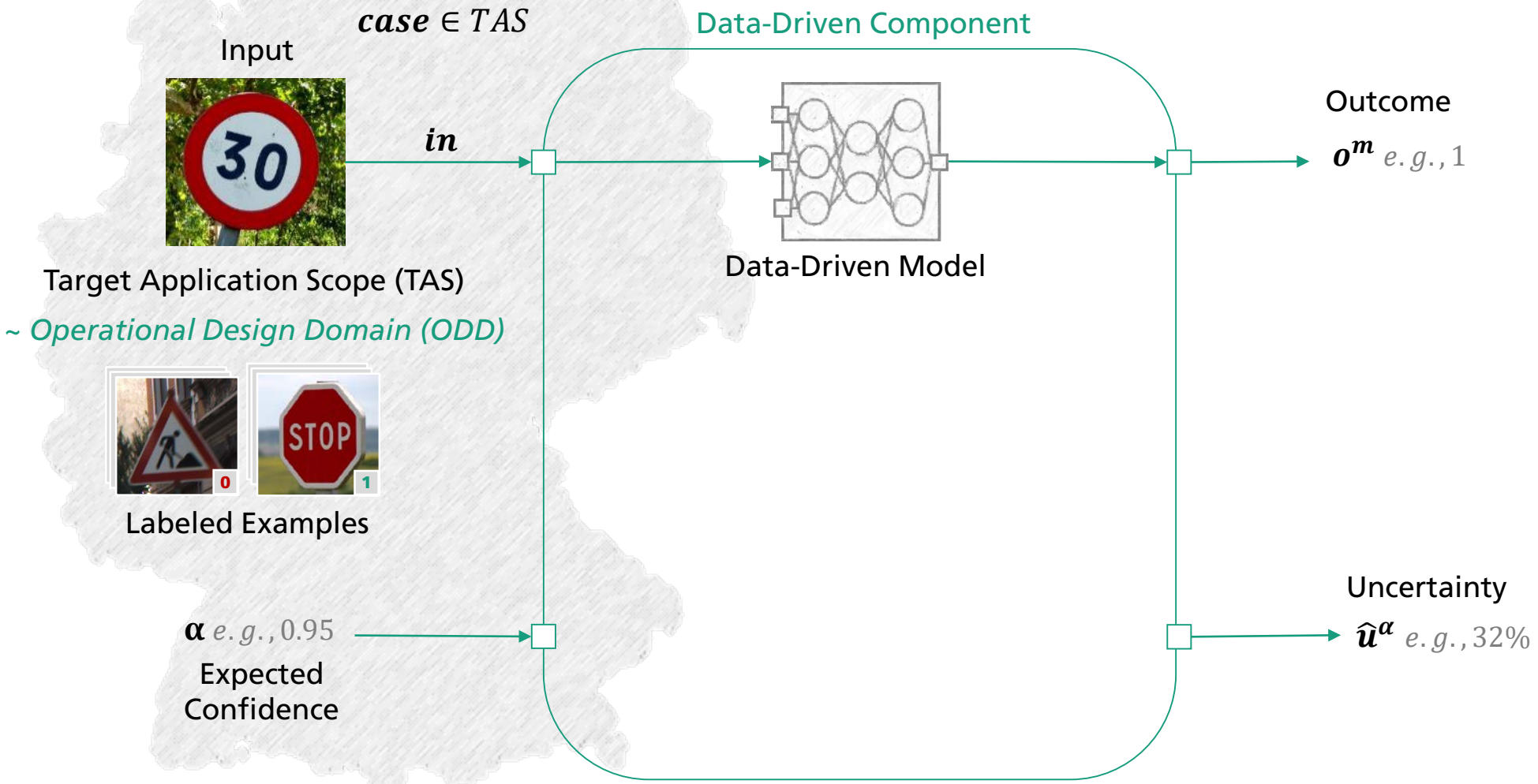


MInD RheinlandPfalz

CrESt

Fraunhofer IESE

Key Sources of Uncertainty: Scope Compliance

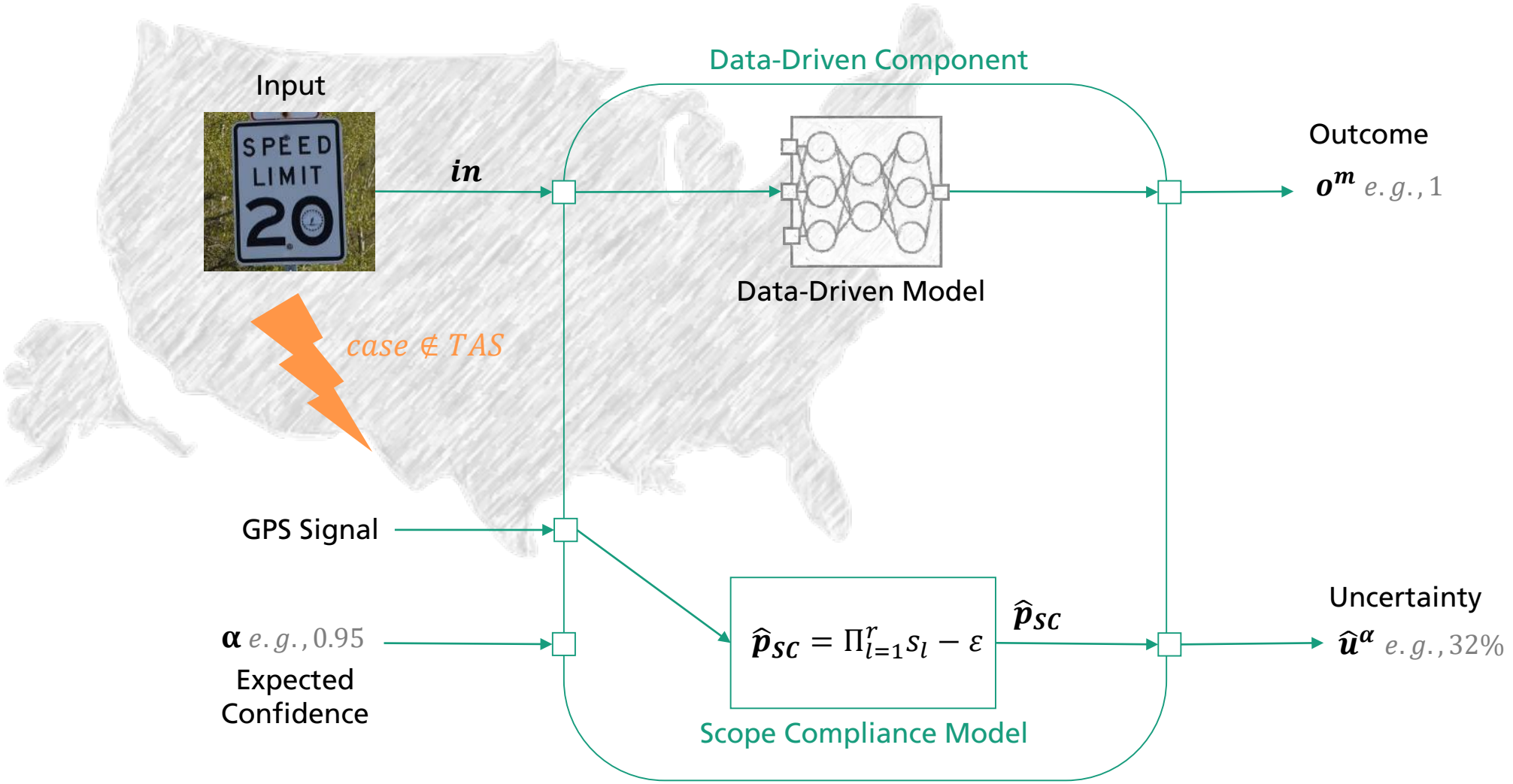


MInD

CrESt

Fraunhofer IESE

Key Sources of Uncertainty: Scope Compliance

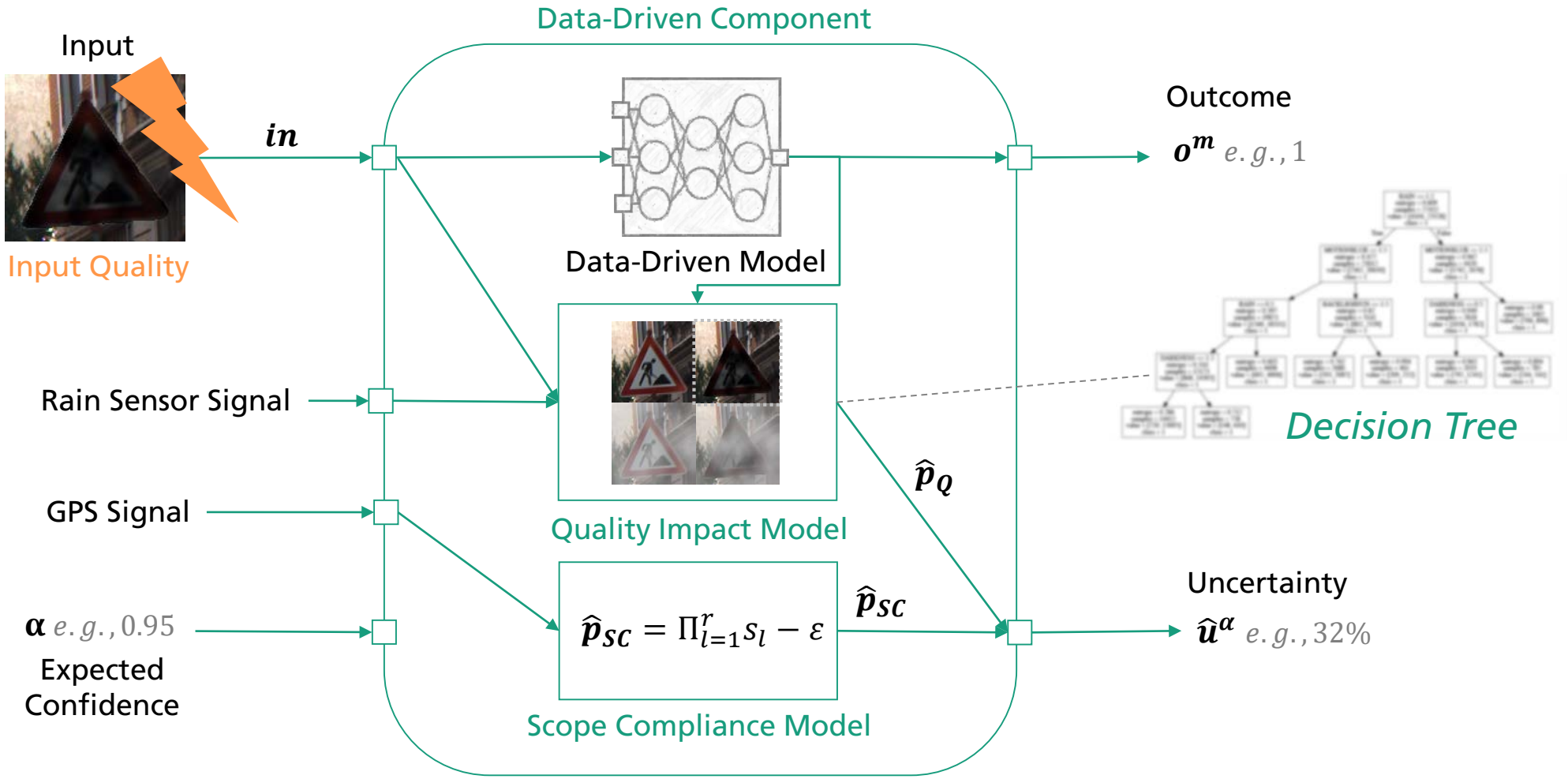


MInD

CrESt

Fraunhofer IESE

Key Sources of Uncertainty: Input Quality

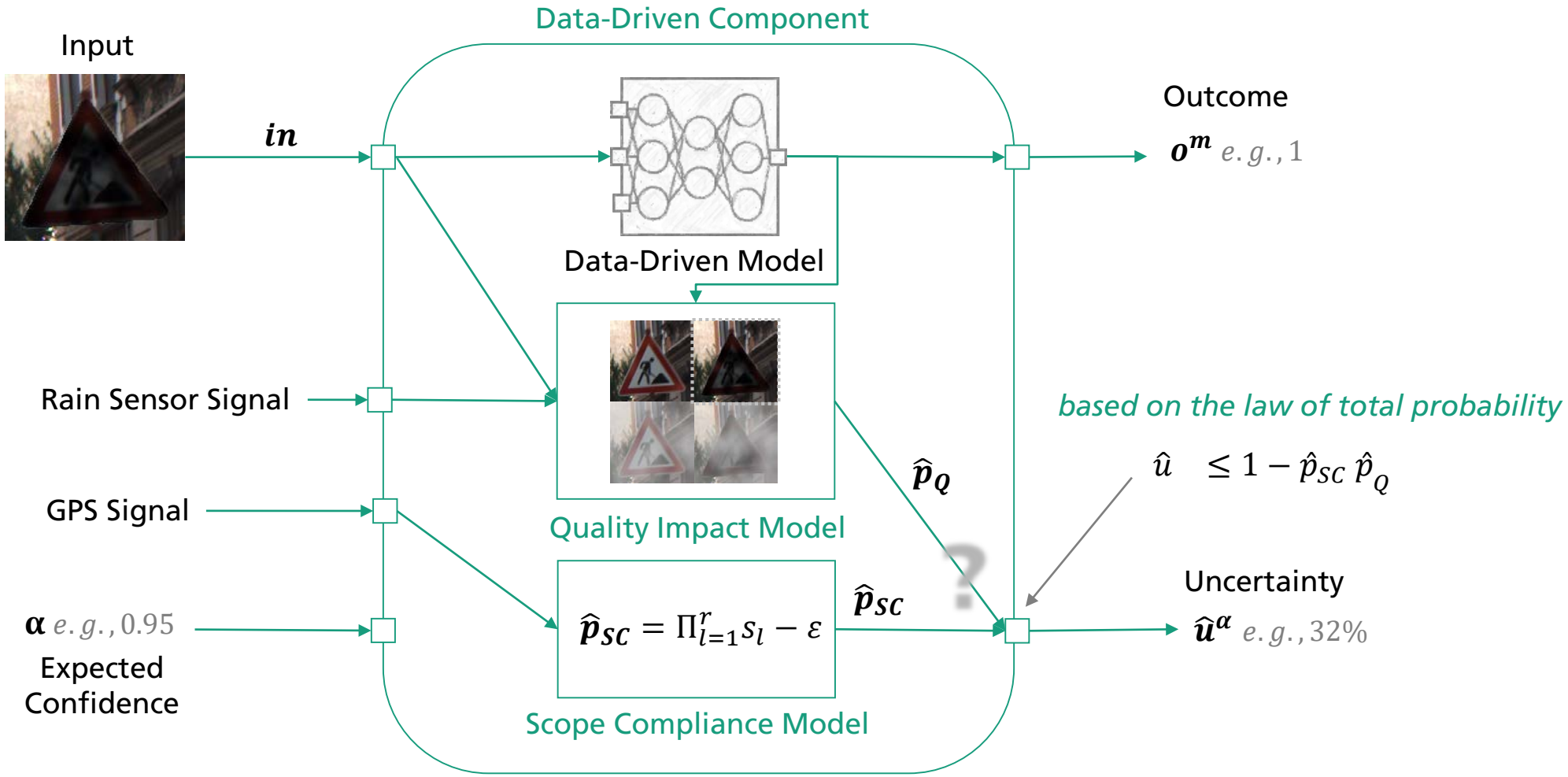


MInD

CrESt

Fraunhofer IESE

Combining Quality Impact and Scope Compliance

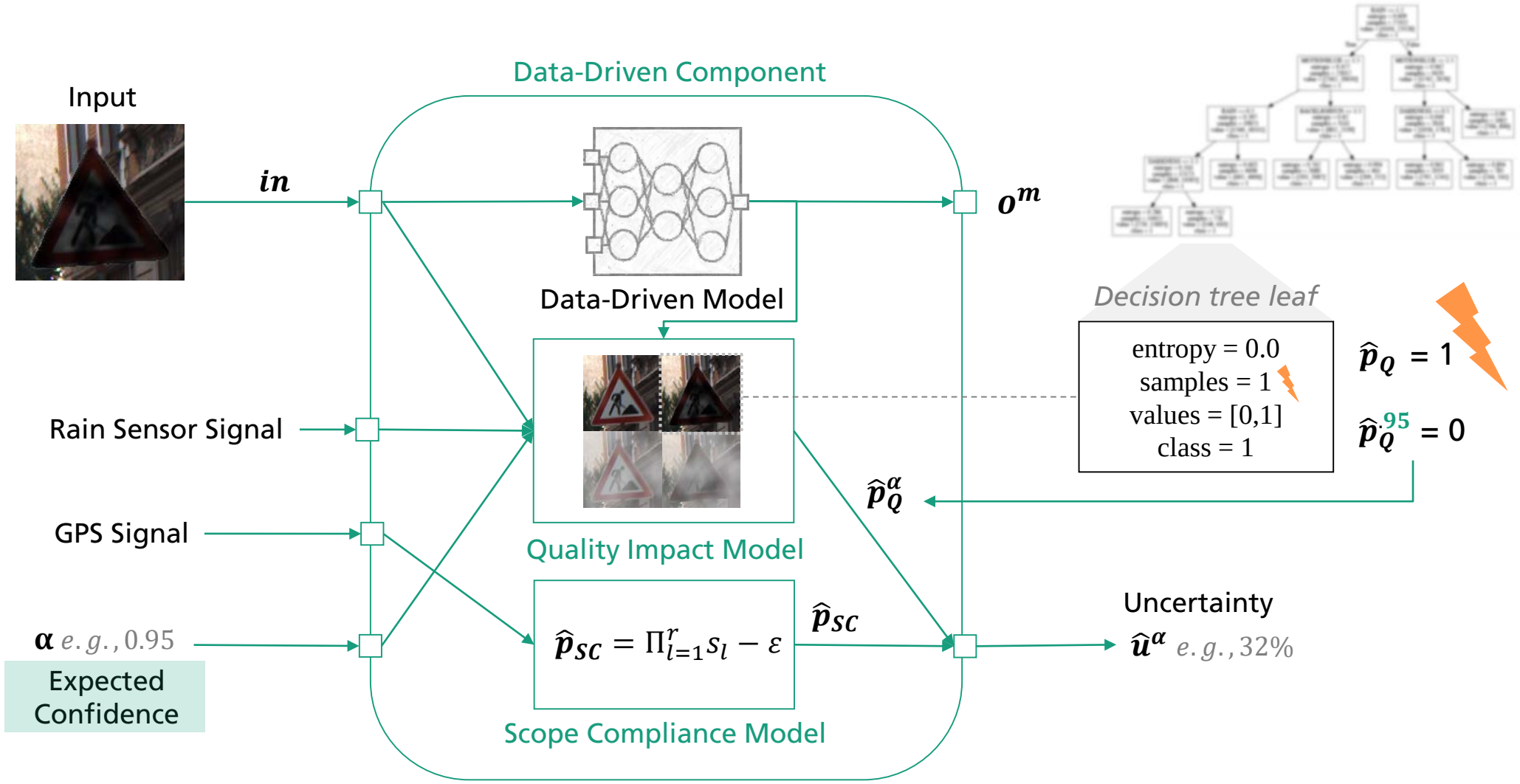


MInD RheinlandPfalz

CrESt

Fraunhofer IESE

Considering Expected Confidence



MInD

CrESt

Fraunhofer IESE

Summary: Key Advantages

Simple and justifiable interpretation based on decision tree structure

Holistic considering model fit, input quality, and scope compliance

Independently applicable from the encapsulated data-driven model

Related References

- Kläs, M., Sembach, L., "**Uncertainty Wrappers for Data-driven Models** - Increase the Transparency of AI/ML-based Models through Enrichment with Dependable Situation-aware Uncertainty Estimates," Second International Workshop on Artificial Intelligence Safety Engineering (WAISE 2019), Turku, Finland, 2019.
- Jöckel, L., Kläs, M., "**Increasing Trust in Data-Driven Model Validation** - A Framework for Probabilistic Augmentation of Images and Meta-Data Generation using Application Scope Characteristics," Proceedings of International Conference on Computer Safety, Reliability and Security (SAFECOMP 2019), Turku, Finland, 2019.
- Kläs, M., Vollmer, A. M., "**Uncertainty in Machine Learning Applications – A Practice-Driven Classification of Uncertainty**," First International Workshop on Artificial Intelligence Safety Engineering (WAISE 2018), Västerås, Sweden, 2018.
- Kläs, M., "**Towards Identifying and Managing Sources of Uncertainty in AI and Machine Learning Models - An Overview**," https://arxiv.org/abs/1811.11669, 2018.

Dr. Michael Kläs

Data Engineering

Fraunhofer IESE, Kaiserslautern

michael.klaes@iese.fraunhofer.de

Acknowledgments. Parts of this work have been funded by the Ministry of Science, Education, and Culture of the State of Rhineland-Palatinate in the context of the MInD project and by the German Federal Ministry of Education and Research (BMBF) under grant number 01IS16043E (CrESt).