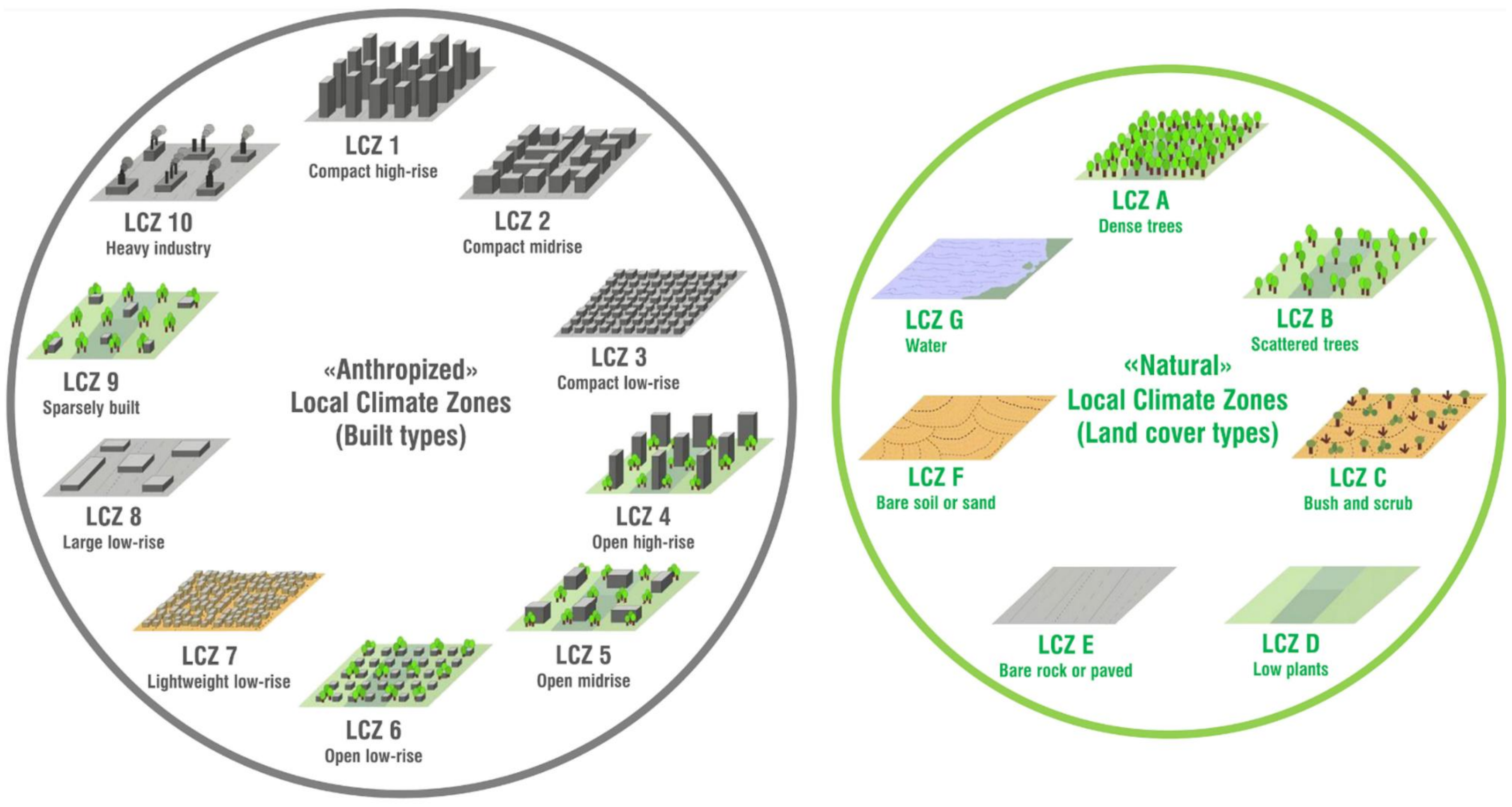
Can Uncertainty Quantification Benefit From Label Embeddings? A Case Study on Local Climate Zone Classification

Christoph Schweden [1], Katharina Hechinger [2], Göran Kauermann [3], and Xiao Xiang Zhu [4,5]

[1]: German Aerospace Center (DLR), [2]: Interhyp, [3]: LMU Munich, [4]:TUM, [5]:Munich Center for Machine Learning



a priori:

$$\alpha \sim p(\alpha|\gamma)$$

a posteriori:

$$f(\alpha|Y) \propto f(Y|\alpha) \cdot f_{prior}(\alpha)$$

$$\text{where } f(Y|\alpha) = \int f(y|\pi) f(\pi|\alpha) d\pi$$

$$\pi|\alpha \sim \text{Dir}(\alpha_1, \dots, \alpha_K)$$

$$\text{Var}(\pi|\alpha) = \frac{\alpha(1-\alpha)}{\sum_k \alpha_k + 1}$$

$$\hat{y}_{conf} = \max_k \hat{\pi}(y = k|x)$$

$$\hat{y} = \underset{k}{\text{argmax}}\, \hat{\pi}(y = k|x)$$

$$\hat{\pi}(y|x) \coloneqq \text{Softmax}(h(x))$$

Top: Local Climate Zone (LCZ) Classification scheme as presented by (Migliari et al. 2024).
Right: Three-level uncertainty pyramid (adapted from (Hüllermeier 2022).

Abstract

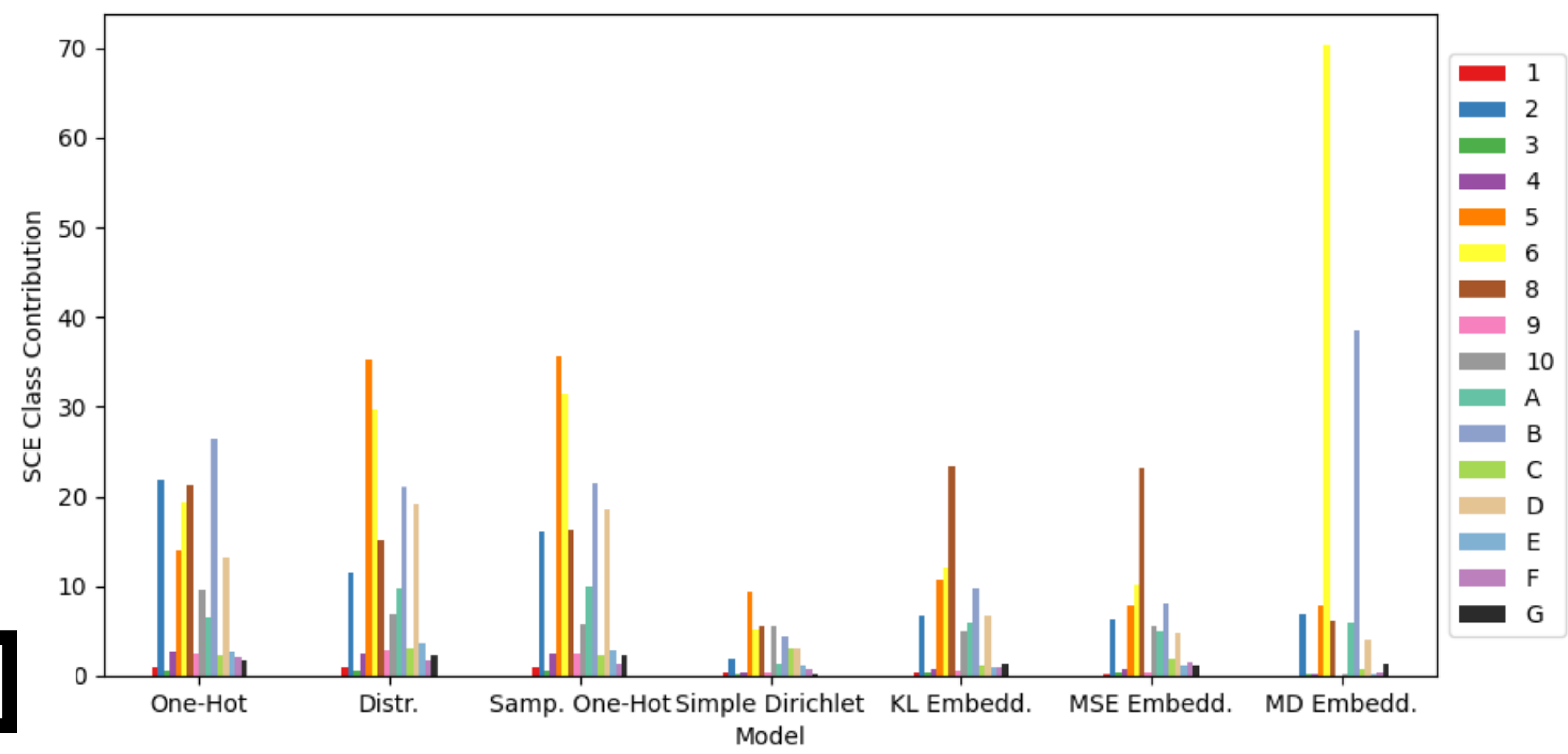
Modern deep learning models have achieved superior performance in almost all fields of remote sensing. An often neglected aspect of these models is the **quantification and evaluation of predictive uncertainties**. A notion of uncertainty indicates the model's **indecisiveness** among the given classes and is essential to understand where the model struggles to classify the data samples

In this work, three levels of uncertainty are distinguished, starting with the typical **softmax pseudo-probabilities** as level-1 uncertainty. As a next level, the more flexible **Dirichlet framework** is utilized as model output space, and hereby also, a **Bayesian setting** with an uninformative prior is considered. For the level-3 uncertainty, an **empirical Bayes** setting is incorporated where a latent embedding of the label space is iteratively estimated by the marginal likelihood of the fully parameterized label space.

The estimated embeddings are then learned by the network in 3 ways: via **two regression losses** and via the **closed-form solution of the Kullback–Leibler (KL) divergence**, utilizing the embedding parameterized as a Dirichlet distribution.
Evaluation is performed with respect to classical performance metrics as well as **calibration** metrics (Table 1). The **static calibration error** (Nixon et al. 2019) is investigated in the Figure below. Have a look at the paper for more results (also OoD detection between built / natural classes).

TABLE I
CALIBRATION AND GENERALIZATION METRICS FOR ALL EVALUATED MODELS. ACCURACY METRICS INCLUDE OA, MAA, WAA, AND KAPPA STATISTIC (κ). CALIBRATION METRICS INCLUDE ECE, MCE, SCE, AND KL DIVERGENCE TO EMBEDDING (KL DIV.). BOLD VALUES INDICATE THE BEST-PERFORMING MODEL FOR EACH METRIC

	OA ↑	MAA ↑	WAA ↑	κ ↑	ECE ↓	MCE ↓	SCE ↓	KL-Div. to Emb. ↓
One-hot	90.3	59.8	88.4	88.4	4.4	18.8	0.6	0.40
Distr.	**91.4**	**80.7**	**91.1**	**89.8**	6.6	23.4	0.6	0.32
Sampled One-hot	90.7	76.7	90.2	89.0	6.8	23.3	0.7	0.32
Simple Dirichlet	89.6	72.4	88.7	87.6	4.1	90.0	**0.2**	0.48
KL Embedd.	89.9	66.9	89.3	88.0	**1.7**	**10.2**	0.3	**0.27**
MSE Embedd.	90.5	72.1	89.9	88.8	2.7	12.6	0.3	0.28
MD Embedd.	73.7	31.3	66.7	67.9	10.2	76.3	0.6	0.83



Main Findings:

- embedding-based approaches show strong performance for calibration
- Competitive results for OoD detection, yet limited by simplicity of the task
- embedded labels offer a flexible framework for incorporating uncertain or ambiguous labels into a supervised training
- Embeddings could be highly beneficial for applications in fields such as urban planning or disaster response.

References:

E. Hüllermeier, "Quantifying aleatoric and epistemic uncertainty in machine learning" in Proc. Uncertainty Artif. Intell., 2022,

M. Migliari, E. Briche, J. Despax, L. Chesne, and O. Baverel, "A cross-analysis matrix comparing multi-site local climate zone trends" Sustain. Futures, vol. 7, 2024

J. Nixon et al., "Measuring calibration in deep learning," in Proc. CVPR workshops, 2019.

Scan for the paper

MUDS Munich School for Data Science HELMHOLTZ · TUM · LMU

LMU

TUM

DLR