

Setting the scene

Traffic accidents emerge from a mix of weather, traffic, and human factors, yet operational datasets typically record only where accidents *have occurred*. This **presence-only nature** limits traditional modeling approaches. We explore **one-class learning** as a realistic way to characterize accident conditions using five years of Dutch accident data enriched with weather and traffic features. By learning directly from observed accidents, we identify recurring environmental patterns linked to increased road-traffic risk. The approach directly supports KNMI's drive toward **impact-based forecasting** (IBF) services, providing building blocks for pilot projects that aim to turn weather and traffic intelligence into operational, impact-based guidance.

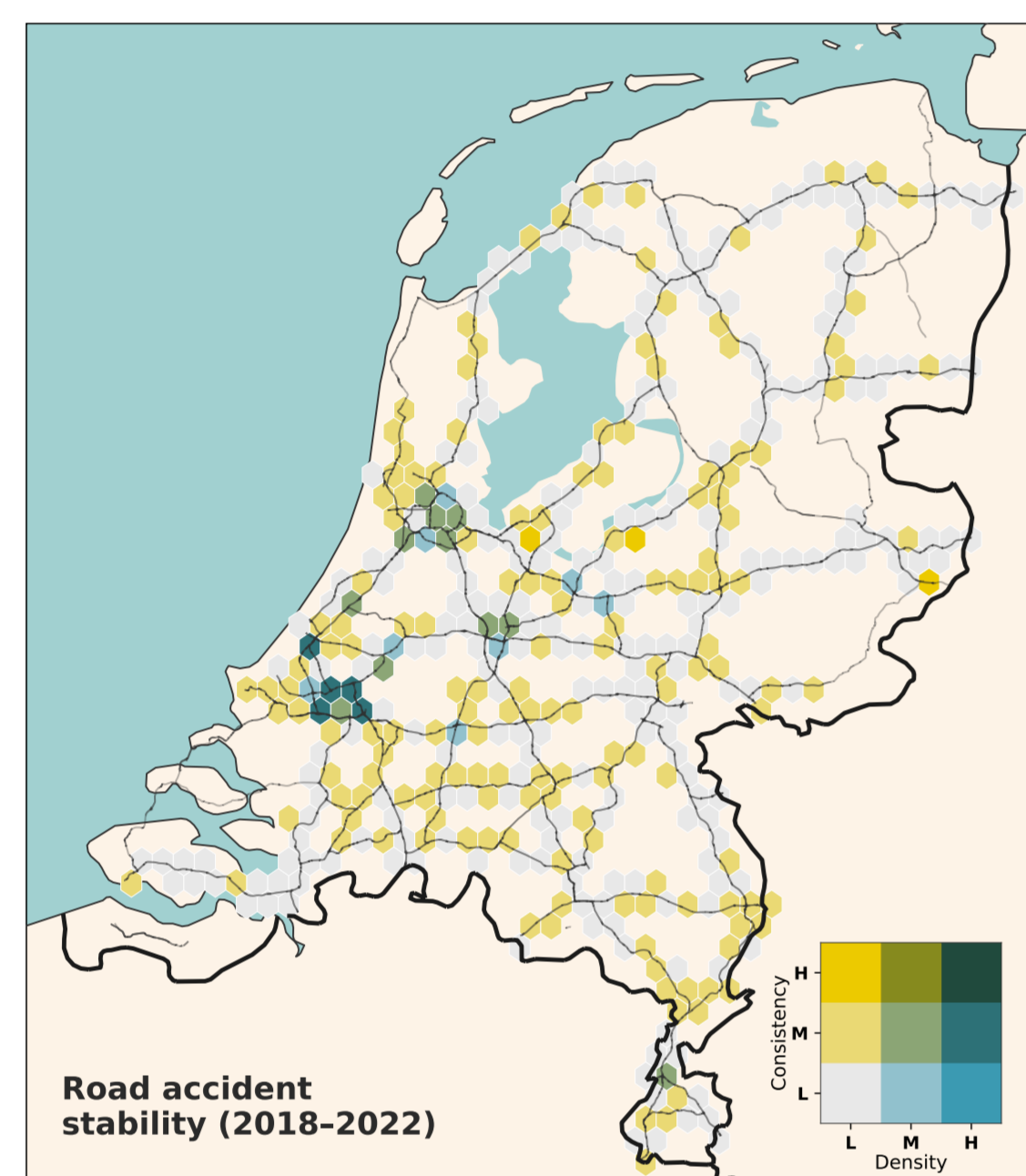


Figure 1. Road accident stability (2018–2022) across the Dutch road network. Bivariate bins highlight the relationship between accident density and temporal consistency (CV).

Modelling road accidents with Deep SVDD

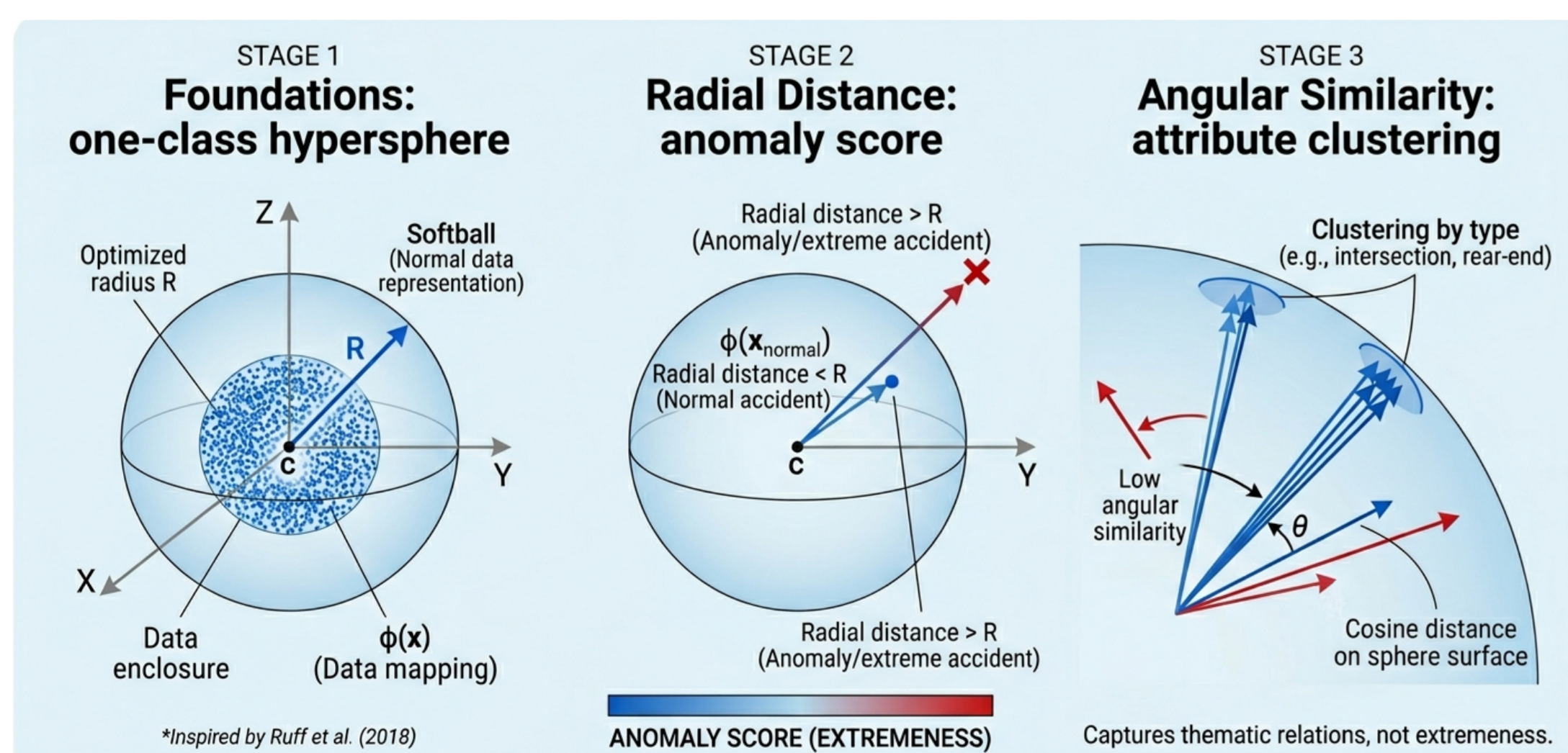


Figure 2. Deep SVDD Conceptual Framework. Road accidents are encoded into a high-dimensional hypersphere (Stage 1), where the learned topology enables dual-path interpretability and explainability: (Stage 2) radial distance as a proxy for anomaly magnitude, and (Stage 3) angular similarity for clustering directional patterns.

Our architecture employs a 3-layer fully connected MLP to project accident data from the Dutch Road Authority (2018–2022) into a 32D latent space. The model is trained on a 60/20/20 split of over 100,000 records of road accidents enriched with weather and traffic features. In this way, Deep SVDD leverages its fully connected structure to capture complex non-linear interactions.

Quantifying accident severity via latent radial distance

Accident embeddings from the 32D latent space are projected onto a unit sphere and analyzed based on their radial distance (R) from the hypersphere center. We categorize the results into 3 interpretable regimes: Conventional ($R \leq 0.15$), representing high-density 'normal' traffic patterns; a Grey Zone ($0.15 < R \leq 0.50$), capturing transitional behaviors; and Extremes ($R > 0.50$), denoting significant statistical anomalies. The seasonal decomposition reveals how environmental factors influence the distribution of these anomalies across the primary Dutch road network.

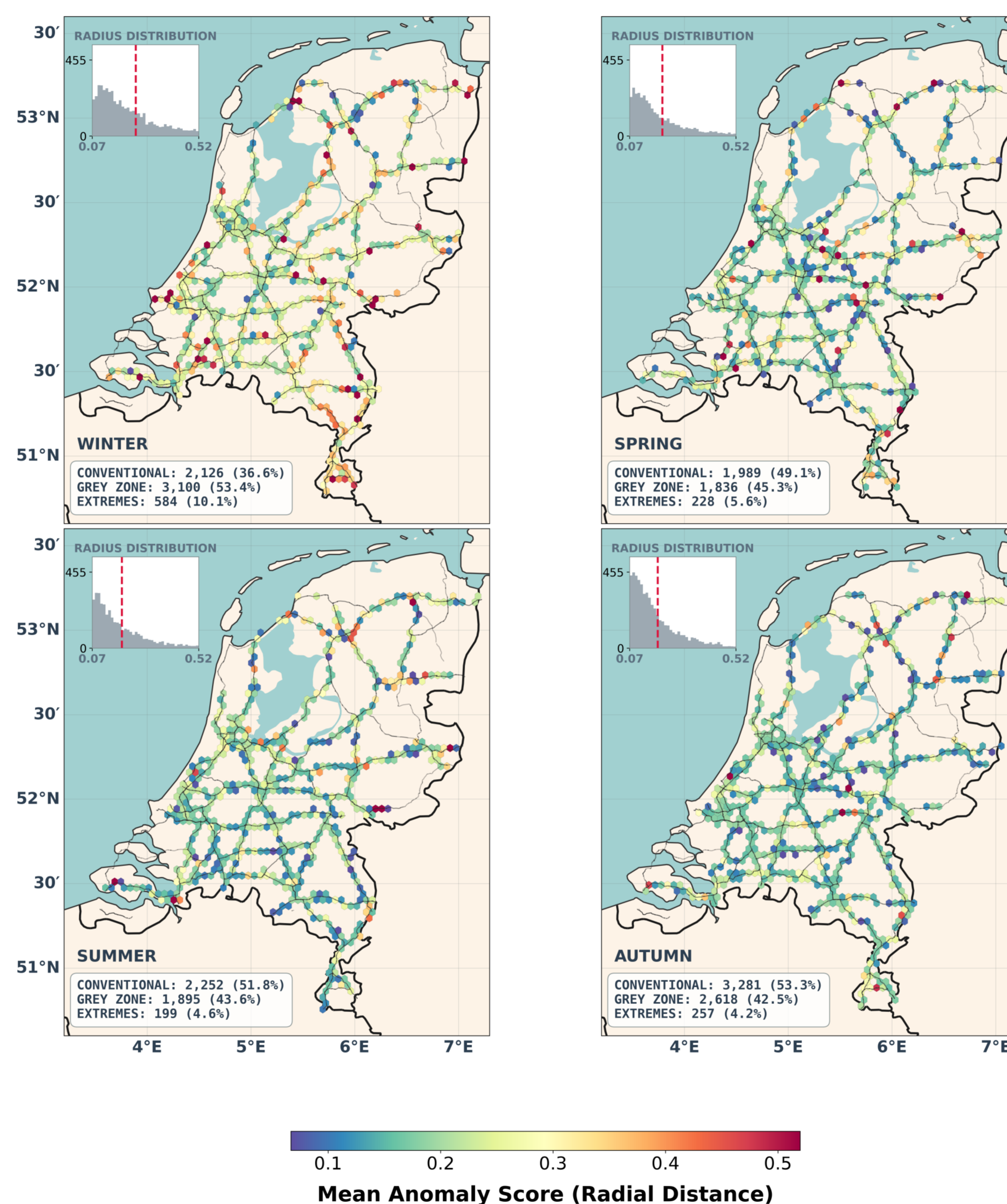


Figure 3. Seasonal distribution of traffic anomalies based on Deep SVDD radial distance (R).

Classifying accident archetypes via angular similarity

To resolve directional weather signatures, we analyzed the 32D latent space via angular similarity. Embeddings were shifted by the hypersphere center and L2-normalized, projecting data onto a unit sphere. A grid search of 30 hyperparameter combinations was conducted, selecting the model with the highest DBCV score and optimal cluster stability. We utilized UMAP for manifold unrolling, mapping the high-dimensional geometry into a 5D space where HDBSCAN could reliably isolate the eight archetypes shown.

#	Variable	Description	N	#	Variable	Description	N
0	mxcap6	Storm Fuel (max, mean, min)	241	4	10fg6, tprate	Wind Gust (mean), Prec. Int. (mean, min)	44,131
1	mxcap6, totalx	Storm Fuel (max, mean), Deep Conv. (max)	339	5	litota6, traffic_int	Lightning Pot (mean, min), Traffic Int.	205
2	mxcap6, tprate, vis	Storm Fuel (max), Prec. Int. (max), Visibility (max)	201	6	mxcap6	Storm Fuel (max, mean, min)	213
3	litota6, cape	Lightning Pot. (max), Atmos. Energy (min, mean)	335	7	litota6, 2t	Lightning Pot. (mean), Temp. 2m (max, mean)	34,978

Table 1. Summary of road accident archetypes

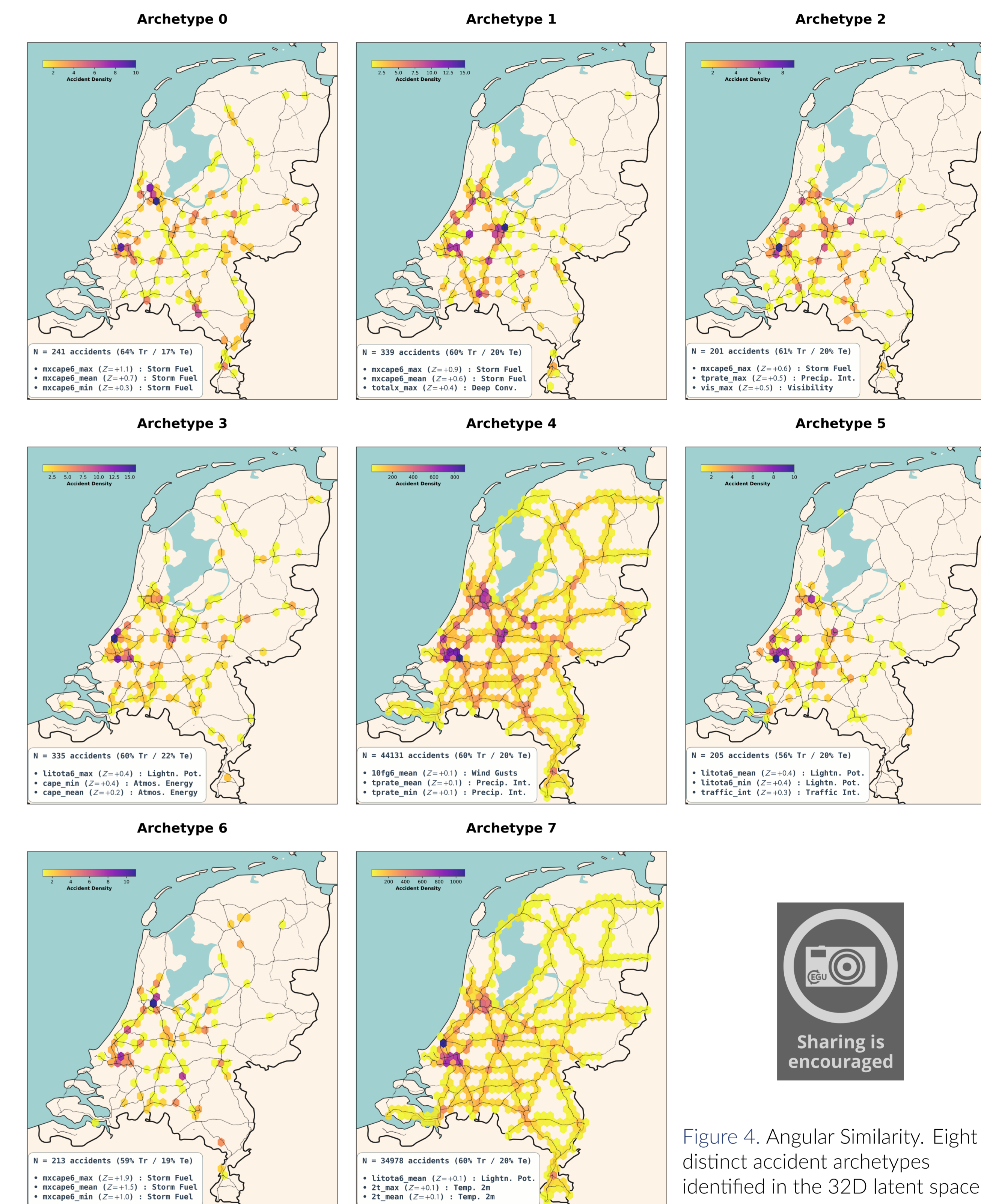


Figure 4. Angular Similarity. Eight distinct accident archetypes identified in the 32D latent space.

Conclusions and future work

We demonstrate that the 32D Deep SVDD latent space provides a robust **explainable geometric framework** for road safety. By mapping accidents onto a high-dimensional hypersphere, we transition from binary anomaly detection to **spatial reasoning** and the discovery of accident archetypes.

Future research will incorporate high-resolution variables—including crowdsourced weather data, solar glare or shadowing effects—to enable **hyperlocal predictions**, while expanding the geometric framework and exploring alternative one-class architectures.