universität freiburg

Exploring Tree-Based Machine Learning Methods for Estimation of Hail Sizes

Amruta Vurakaranam[1], Christian Berndt[2], Katharina Lengfeld[2], Markus Schultze[2], Katharina Schröer[1]
[1]Albert-Ludwigs-Universität Freiburg, Germany
[2]Deutscher Wetterdienst, Germany

1. Hailstorm Analysis, Impacts, and Prediction Initiative (HAIPI)

- Aims to improve estimation of hail, particularly hailstone size in Germany
- Radar-based algorithms exist to detect hail, but they can still fail to accurately identify hail-producing storms[1]
- Integrating new data sources like crowd-sourced reports and dual-polarization radar

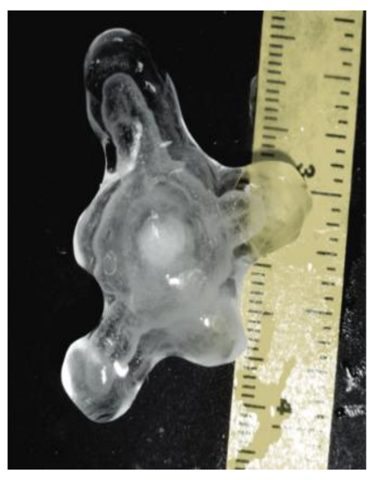
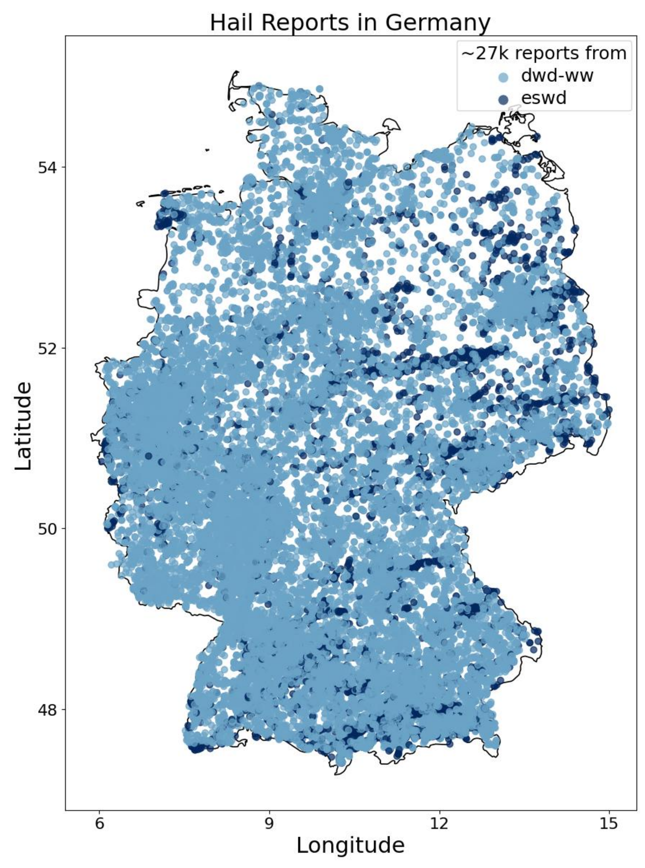


Photo by H. Pogorzelski.

2. Datasets – Crowdsourced Reports + Radar

Crowdsourced Reports

- Have size information
- European Severe Weather Database (ESWD)[2] – 4k
- DWD WarnWetter[3] – 23k

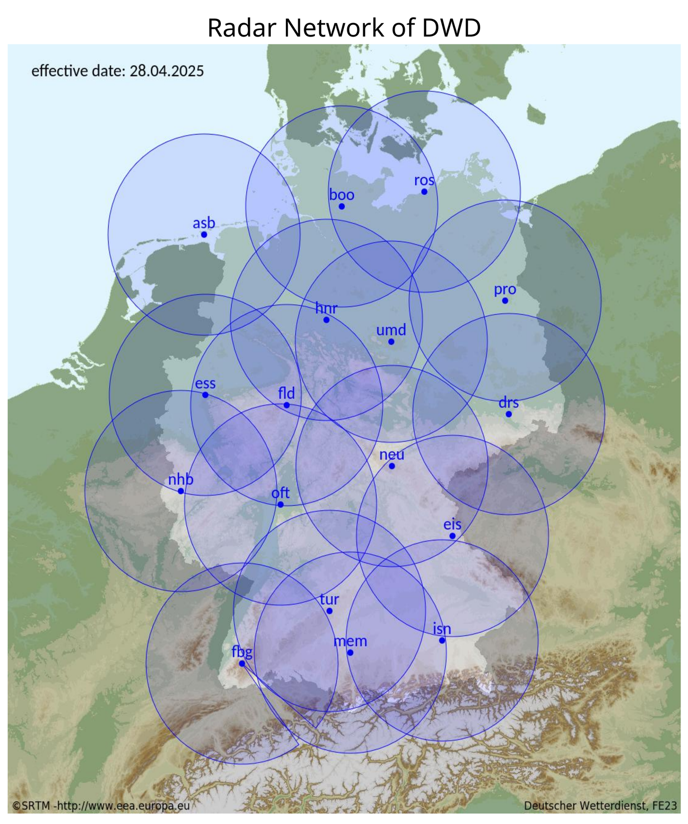


Quality Control of Reports

- ESWD – QC1 and QC2
- DWD WarnWetter – adapted from Barras et al 2019

Radar Network[4]

- 17 doppler radars from DWD
- Upgraded to dual-pol



KONRAD3D[5]

- Storm tracking algorithm; provides storm objects
- Radar + lightning + hydrometeor classification
- Every 5 mins

3. Machine Learning Experiments

- Links KONRAD3D cells to hail reports (2016–2025)
- Each storm cell is linked to all associated hail reports; report with the largest hail size is retained
- Utilizes KONRAD3D parameters (e.g., vertical extent, echotop) as features
- Predicts hail size (target) using multi-class classification

Class	Class Name	Size
0	Graupel/Hail	0.5 – <1cm
1	Small Hail	1 – 2cm
2	Medium Hail	2 – <5cm
3	Large Hail	>= 5cm

- Start with tree-based models and then move on to more complex models like CNNs
- Use different importance scoring for feature selection and understanding the models

4. Do simple tree-based models work?

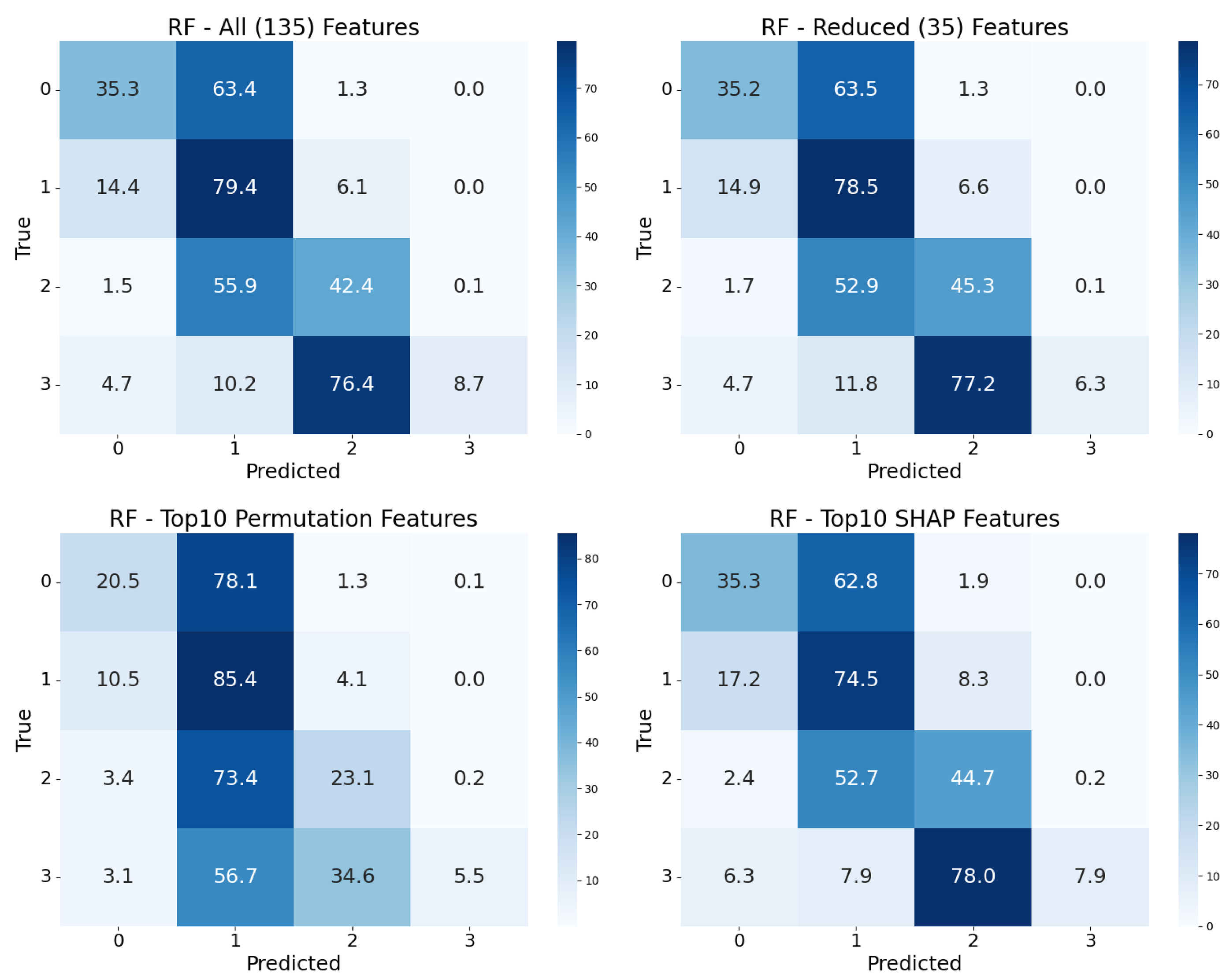
- Random Forest (RF), XGBoost (XGB), and CatBoost (CB) evaluated with 5-fold Stratified Group K-Fold
- **Capture common hail sizes well but struggle with rare, large-hail events**.

Model	Accuracy	Balanced Accuracy	F1 (weighted)	F1 (macro)	ROC AUC
RF	0.64 ± 0.01	0.41 ± 0.03	0.62 ± 0.01	0.44 ± 0.05	0.78 ± 0.03
XGB	0.65 ± 0.01	0.38 ± 0.03	0.61 ± 0.01	0.41 ± 0.05	0.78 ± 0.02
CB	0.65 ± 0.01	0.37 ± 0.02	0.61 ± 0.01	0.39 ± 0.02	0.79 ± 0.02

5. Which features drive model predictions?

- Feature reduction: hierarchical clustering + representative feature (using highest mean correlation)
- Stability analysis (1000 resampled runs) → stable top-10 predictive features using permutation importance
- SHAP (SHapley Additive exPlanations) on all reduced features
- Case studies with all features vs reduced/top-10

→ Can the models be explained?



- **Reduced features retain most predictive performance of full model**
- SHAP top-10 features are sufficient to explain model predictions across most hail classes
- Permutation top-10 features emphasize global accuracy, slightly underperforming on rare events

6. Outlook

- **Simple models work:** Storm structure (e.g., echo geometry, hail area), intensity measures (e.g., dBZ thresholds, volume above thresholds), and environmental predictors (e.g., wind shear, NWP-derived parameters) drive model predictions
- Next steps: CNNs to exploit the 3D spatial and polarimetric characteristics of dual-pol radar data

[1]Barras, H., Hering, A., Martynov, A., Noti, P.A., Germann, U. and Martius, O., 2019. Experiences with> 50,000 crowdsourced hail reports in Switzerland. *Bulletin of the American Meteorological Society*, *100*(8), pp.1429-1440.
[2]Dotzek, N., P. Groenemeijer, B. Feuerstein, and A. M. Holzer, 2009a: Overview of ESSL's severe convective storms research using the European Severe Weather Database ESWD. Atmos. Res., 93, 575−586
[3,4,5]From DWD Open Data

Amruta Vurakaranam
amruta.vurakaranam@geographie.uni-freiburg.de
Climate Geography Lab Freiburg

HAIPI

Grant 4823EMF05

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

DWD