Evaluating Image-Derived Features for Mountain Forest Classification from Historical KH-9 Imagery

Roman Sitko

Faculty of Forestry, Technical University in Zvolen

RETROCHANGE – USE CASE

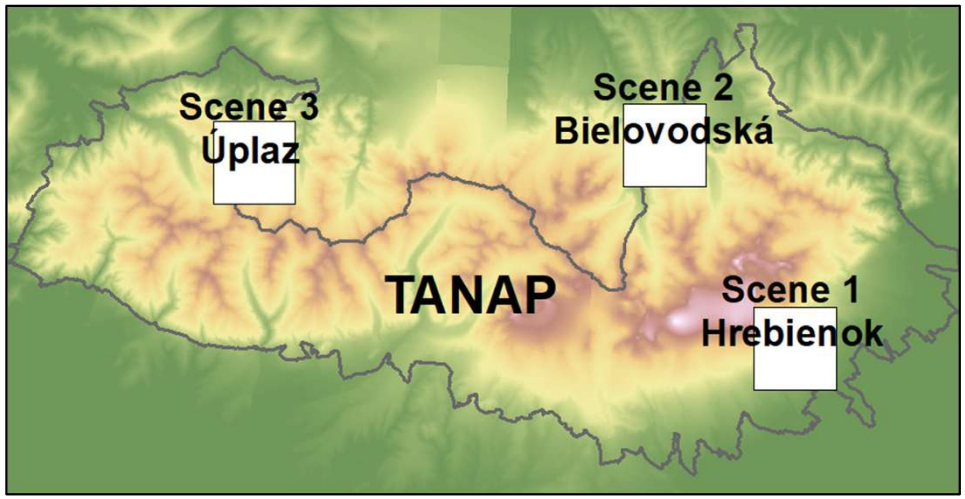


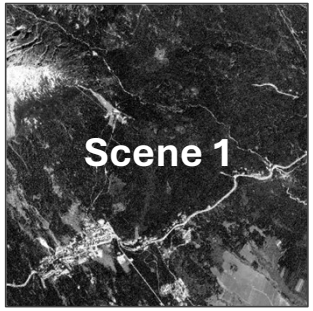


Motivation: historical mapping in a difficult mountain setting

KH-9 imagery provides fine spatial detail, but its interpretation is constrained by a single panchromatic band and variable imaging conditions.

3 scenes (5x5 km, 0,5 m pixel) – acquisition time Jun 27th 1975









Challenge

- single panchromatic band: no multispectral vegetation indices such as NDVI
- High Tatras: pronounced terrain shadows and illumination differences
- forest, dwarf pine/shrubland and other land cover can have overlapping panchromatic intensity distributions
- transfer between KH-9 study areas remains uncertain

Core questions

Which feature families improve classification?

Which ones actually transfer to a held-out KH-9 scene?

Experimental design and validation

Feature selection uses Scenes 1–2. Scene 3 labels are held out for evaluation, local labels enter only the adaptation branch.

Model development: Scenes 1–2 (Hrebienok + Bielovodská)

Scenes 1–2 reference polygons — 43 image-derived features — **Random Forest** fixed settings — Spatial CV 3 folds — Feature-family screening — **Selected features** — **Final transfer RF trained on Scenes 1–2**

Cross-scene transfer: held-out Scene 3 (Úplaz)

Predict Scene 3 — **Independent test metrics** — Map output GeoTIFF

Independent test metrics ← Scene 3 reference CORE + conf=1

Scene-adapted model: Scene 3 samples

Scene 3 local training folds — **Adapted RF with local samples** — Spatial-CV metrics within Scene 3

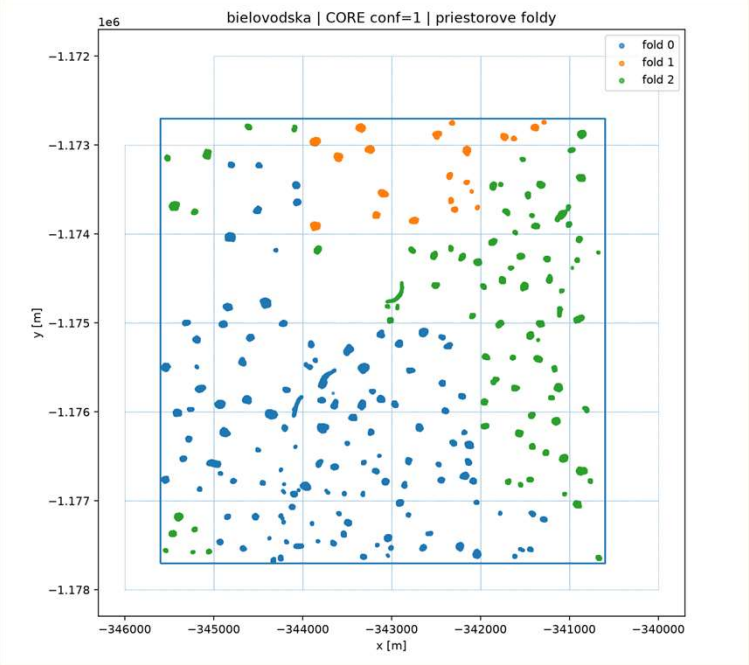
Used to diagnose the value of local reference samples, not to choose the transfer model.

Target classes and reference data

Reference polygons are interpreted visually, subclass codes remain available for error analysis.

Final mapped classes

- 1 Forest
- 2 Dwarf pine and other shrubs
- 3 Other land cover (non-forest)



Table

uc1_3_5x5_uplaz_training

FID	Shape *	Id	subclass	main_cls	site_id	confidence	train_role
25	Polygon	29	34	3		1	core
26	Polygon	28	34	3	L13	1	core
27	Polygon	27	34	3	L13	1	core
38	Polygon	39	34	3	L13	1	core
39	Polygon	40	34	3		1	core
69	Polygon	70	34	3		1	core
101	Polygon	102	34	3	L12	1	core
102	Polygon	103	34	3		1	core
131	Polygon	132	34	3		1	core
132	Polygon	133	34	3		1	core
133	Polygon	134	34	3	L06	1	core
134	Polygon	135	34	3	L06	1	core
178	Polygon	179	34	3	L01	1	core
179	Polygon	180	34	3	L01	1	core
180	Polygon	181	34	3	L01	1	core
181	Polygon	182	34	3	L02	1	core
182	Polygon	183	34	3	L02	1	core
183	Polygon	184	34	3	L03	1	core
184	Polygon	185	34	3	L03	1	core
185	Polygon	186	34	3		1	core
28	Polygon	25	35	3		1	hard
29	Polygon	26	35	3		1	hard
22	Polygon	23	36	3		1	core

1 (0 out of 255 Selected)

uc1_3_5x5_uplaz_training

Reference rules

- RF model fitting: CORE + confidence = 1
- polygons are eroded by 7 m to reduce boundary mixing
- confidence = certainty of reference interpretation
- quality control (QC) mask excludes uninterpretable centres
- site_id and cv_group keep related locations together

Reference data for evaluation on Scene 3

*** 47 538 evaluation samples ***
CORE + conf=1, after QC and exclusion of invalid feature values

Feature families tested

Full-raster image features were generated first, terrain features were sampled only where DEM coverage exists.

baseline

DN

1 feature

Original 8-bit brightness

Local statistics

9 features

Mean, standard deviation, Shannon entropy
5×5 / 11×11 / 21×21

GLCM

9 features

Contrast, homogeneity, ASM
3 windows, 16 grey levels, 4 directions

PATTERN

6 features

NDC, CVN
from imagery quantised to 16 grey levels

Extended texture

18 features

Gabor, morphology, LBP
local texture descriptors

Terrain

4 features

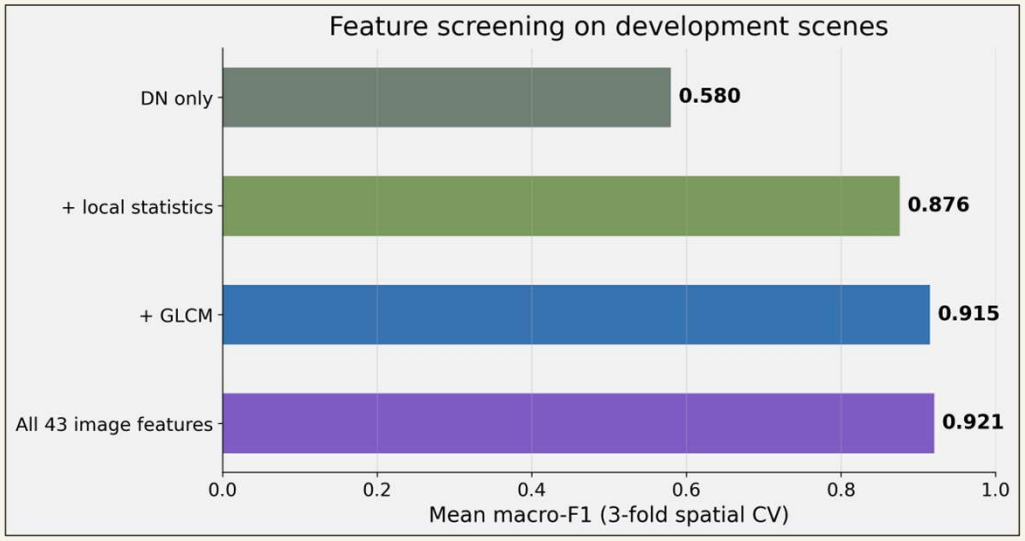
Elevation, slope, sin/cos aspect
DEM-valid samples only

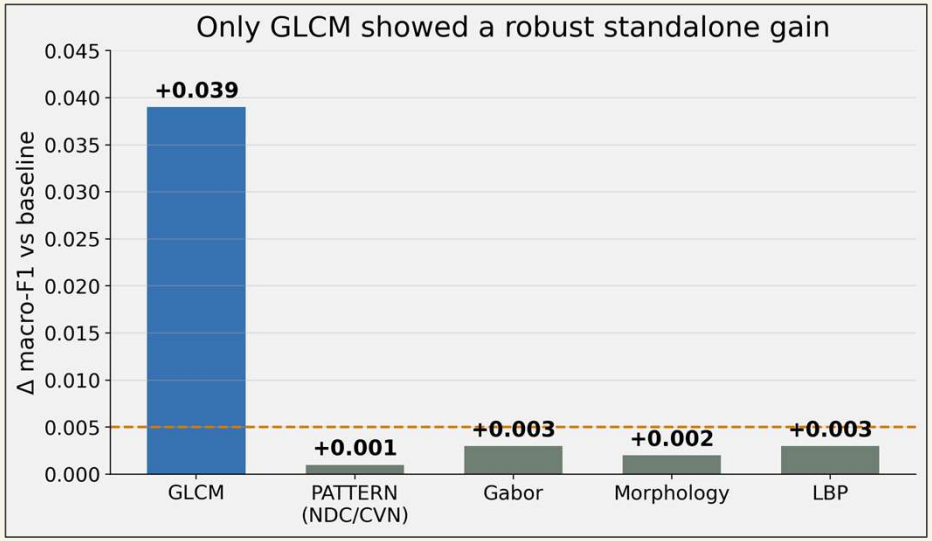
Retention rule: a feature family must improve spatial-CV performance, visual attractiveness alone is not sufficient.

Feature screening: local context and GLCM matter most

Development scenes 1-2: Hrebienok + Bielovodská, 3-fold spatial CV.

Macro F1 = average class-wise balance between precision and recall

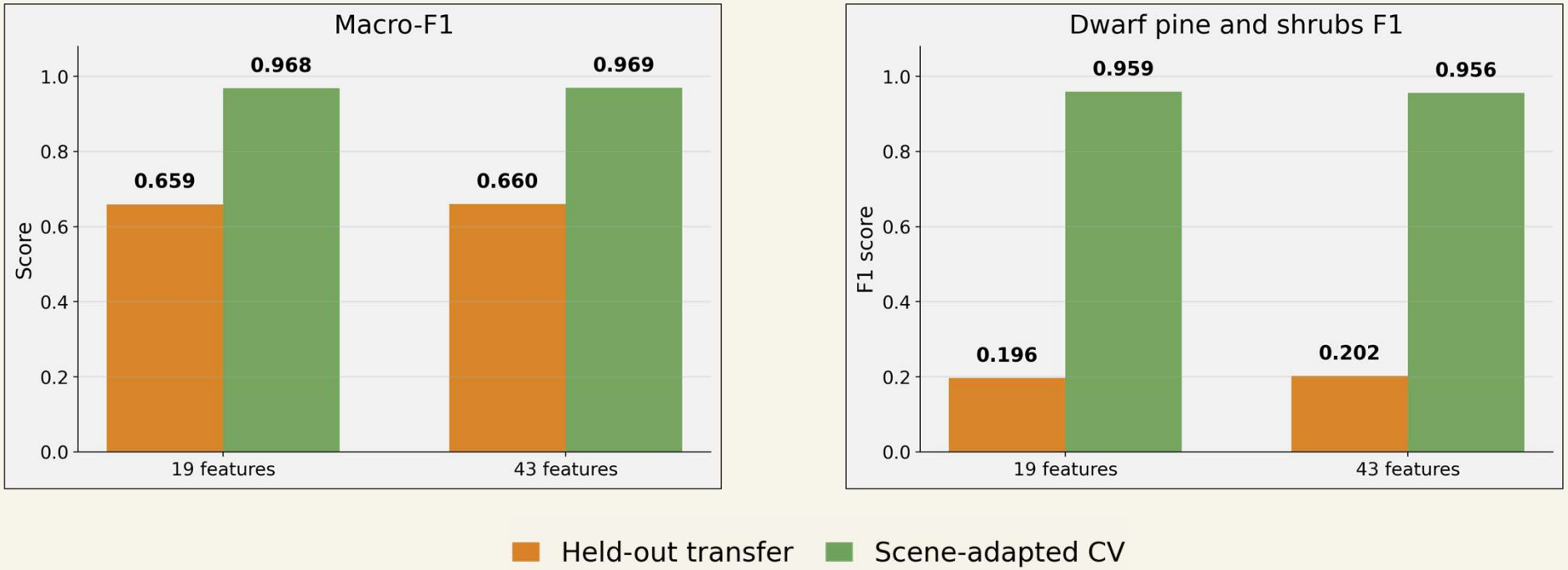




Interpretation: Local statistics produced the largest gain. GLCM was the only individually added family to exceed the retention threshold and the full 43-feature set was only marginally higher than the selected 19-feature set.

Transfer to held-out Scene 3

Held-out transfer and scene-adapted spatial CV represent different deployment settings.

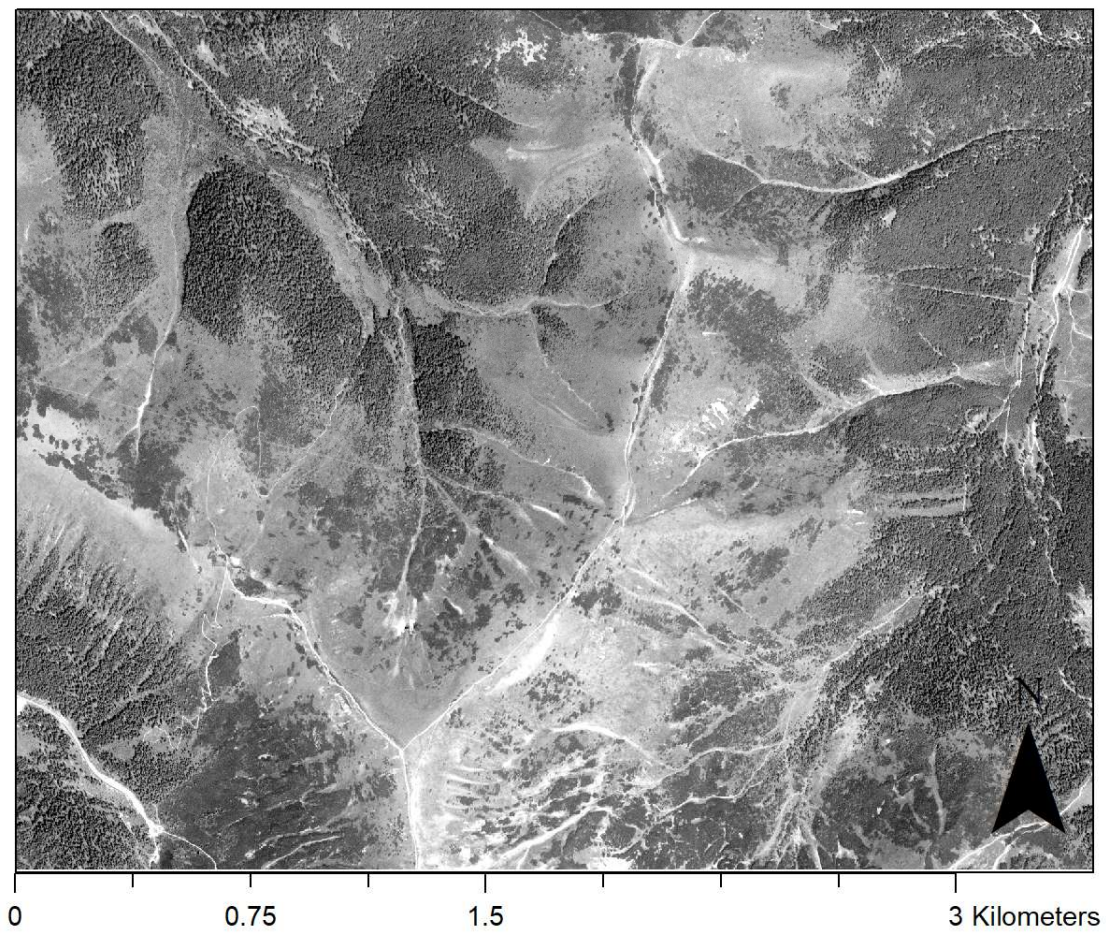


Transfer test: model trained only on Scenes 1–2.
Scene-adapted CV: training includes local Scene 3 folds and validates on held-out Scene 3 folds.

Key message: transfer to Scene 3 missed most dwarf pine and other shrubs. Local Scene 3 samples largely solved the class-2 problem under spatial CV.

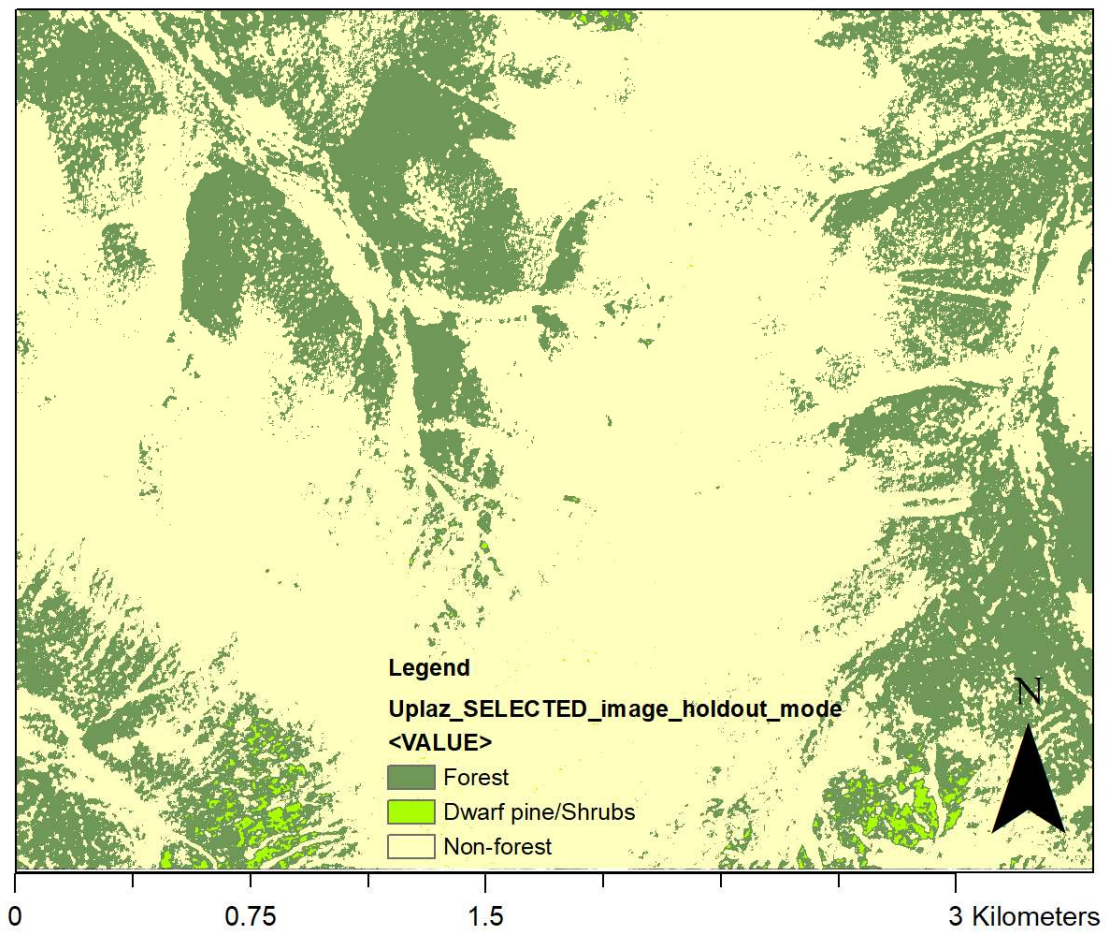
Visual comparison: original KH-9 image – Scene 3

KH-9 → transfer RF → scene-adapted RF.



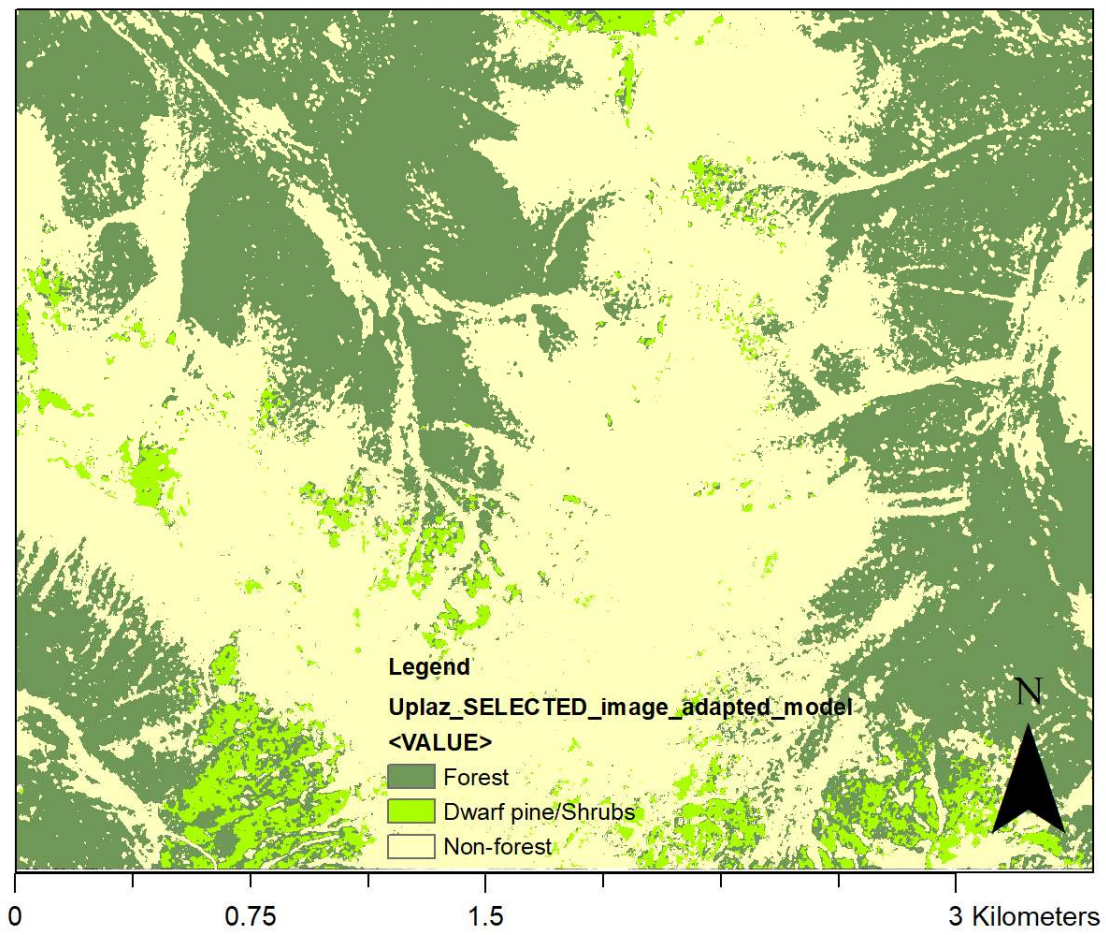
Cross-scene transfer scenario: conservative class 2 prediction

Model trained on Hrebienok + Bielovodská only, evaluated on Uplaz.



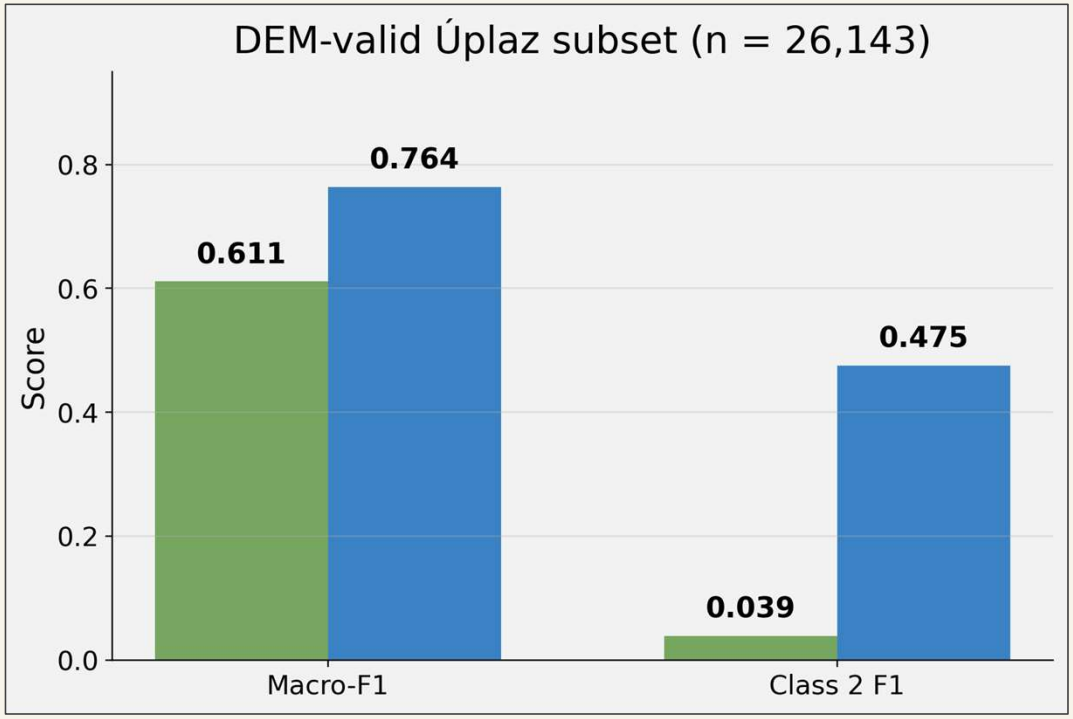
Scene-adapted scenario: local samples unlock the Scene 3 class structure

Development data plus Scene 3 spatial folds, not an independent transfer result.



Terrain features: DEM improved gain of cross-scene transfer

DEM is incomplete, the terrain comparison below uses identical Scene 3 samples with valid DEM values.



Matched-sample DEM effect

Image only Image + terrain

- same 26 143 DEM-valid Scene 3 samples
- macro-F1: 0.611 → 0.764
- class 2 F1: 0.039 → 0.475
- class 2 recall: 0.020 → 0.311

Interpretation

DEM adds ecological context (elevation, slope and aspect). It should remain an optional mountain workflow module, evaluated only where DEM coverage exists.

Main conclusions

What the experiment tells us about building KH-9 forest maps.

1 Local image context mattered most

DN alone was insufficient; local statistics and GLCM carried most of the useful signal in this experiment.

2 GLCM exceeded the retention threshold

It was the only individually added family to exceed the feature-retention threshold.

3 The compact 19-feature set is defensible

All 43 image features gave almost the same transfer score as the selected 19 features.

4 Transfer was the hard problem

The transfer RF missed most Úplaz dwarf pine/shrub samples, while classes 1 and 3 performed much better.

5 Local reference samples changed the picture

Under the evaluated local-adaptation protocol, scores were much higher than transfer-only RF.

6 DEM is an optional mountain module

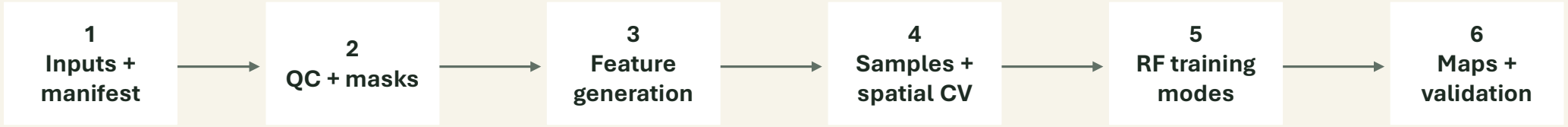
Terrain improved class 2 in transfer on DEM-valid samples, but should remain optional and tested separately.

Practical recommendation: final workflow should support both transfer-only and scene-adapted modes, with terrain as an optional module.

Proposed reproducible workflow

The same validation module can evaluate RF outputs and external classification maps under matched class definitions and masks.

Workflow as a set of modules (JupyterLab notebooks)



Two operational modes

Transfer-only mode
Train on available scenes, test on a new scene. Report transfer performance explicitly.

Scene-adapted mode
Add local target-scene samples and validate with spatial folds.

Evaluation of external classification maps: harmonise class definitions, masks, CRS/resolution and sampling points, then compute the same metrics.

Thank you for your attention

Questions and discussion

From KH-9 to Forest Maps





Roman Sitko
sitko@is.tuzvo.sk