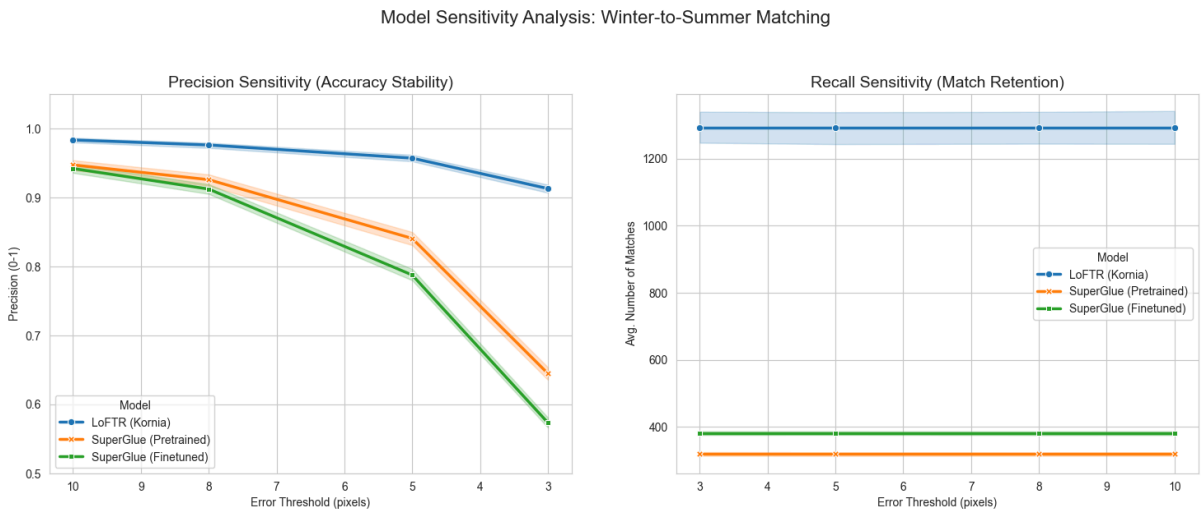
Performance

Task 1: Mountain NER

Context	precision	recall	f1-score
Overall	0.92	0.91	0.91
Ukrainian	0.88	0.88	0.88
English	0.95	0.92	0.93

Task 2: Sentinel-2 image matching



Future improvements

Note: general reviews of the tasks are written in their README files. Here we outline just future improvements.

General (applies to both tasks)

1. **Hyperparameter tuning**

- Why: Current results are promising but sub-optimal.

- How: run searches (grid / random) over key parameters.

2. **Expand dataset**

- Why: More data reduces overfitting and teaches models edge-case behaviors.

- How:
    - For NER: add negative samples where tokens look like mountains (brand names, clubs, bottles) and contrasting contexts;
    - For satellites: add more tiles, seasons, atmospheric conditions, and imaging years.

Task 1 - Mountain NER

1. **Inject context-contradictory samples**

- Why: NER fails in such samples:

Футбольний клуб "[Карпати]" зіграв унічию сьогодні.

- Solution: Expand dataset with counter example samples.
- How: add an LLM prompt to specifically generate *confusing negatives* and integrate them into training; include token-level alignment checks.

2. **Active sampling & human-in-the-loop**

- Why: The model will make consistent errors when using the generated dataset.
- How: Collect high-confidence false positives & high-confidence false negatives with human dataset labeling

---

Task 2 - Sentinel-2 image matching

1. **Solve variable-size keypoint batching**
- Why: Per-sample forward pass (no parallelism) is a major bottleneck.
- Solutions (potential):
    - **Padding + Masking**
    - **Bucketing**

2. **Address fine-tuned model jitter**

- Possible causes:
    - Overfitting to seasonal textural features.
    - Learning dynamics (too high lr or unstable gradients).

- Approaches (possible):
    - **Temperature / confidence regularization**
    - **Smoothing / sub-pixel refinement**

3. **Data & augmentation**
   Add more season pairs, atmospheric/cloud variations, sensor noise simulation

---