Spatiotemporal Inference of Landslide Events by Fusing Geographic Features and Social Media Information

Bing Xu[1,*], Alexander Brenning[1]

[1]Department of Geography, Friedrich Schiller University Jena, Loebdergraben 32, 07743 Jena, Germany

Abstract

Social media posts can provide rapid clues about landslides, but location mentions are often vague and the event time may differ from the posting time. Point-based geocoding cannot represent within-footprint uncertainty or support evidence fusion. We propose an uncertainty-aware spatiotemporal inference framework that generates a landslide likelihood map within the footprint returned by toponym resolution and estimates a plausible trigger date. Within the footprint, we fuse a susceptibility prior with proximity evidence from roads and water bodies, whose weights are modulated by CLIP-based zero-shot image cues, and we further refine scores using PRISM daily precipitation. We evaluate maps using manually annotated post locations and assess ranking quality and search-space reduction via percentile rank (PR), hit rates, and `Area@Top`k%. The results show improved ranking metrics and markedly smaller high-probability regions compared with a susceptibility-only setting.

Keywords

Geographic Information Extraction, Spatiotemporal Inference, Large Language Model, Uncertainty-aware Geocoding

1. Introduction

Social media can support rapid situational awareness and emergency response because people often post text and photos soon after a disaster, offering clues about what happened, where, and when [1, 2, 3]. For landslides and related hazards, however, location cues in text are often vague and coarse, such as an administrative area, a road name, or a regional label [4]. Posting time can also differ from the actual time of the event [5]. These issues limit standard point-based geocoding: forcing a place name to a single coordinate ignores spatial extent and within-area uncertainty, which can bias downstream assessment and response decisions.

Toponym resolution and disambiguation do not require the output to be a single point. A more faithful approach maps a place mention to its geographic footprint, such as a point, line, or polygon, which better captures differences in spatial scale and explains ambiguity [6]. However, for disaster localization, a footprint alone still leaves a critical question unanswered: where within the footprint is the event most likely to have occurred? Prior work in geographic information extraction typically follows a pipeline of place-name recognition, candidate generation/disambiguation, and spatial representation [7], but bridging footprints to actionable event locations requires integrating external evidence within the candidate area, such as terrain, proximity to infrastructure and hydrologic features, and triggering signals like rainfall [8]. Large language models and geography-guided prompting have also been explored to improve robustness and interpretability in place-name recognition and disambiguation [9, 10]. Meanwhile, vision–language models make it possible to extract interpretable scene cues (e.g., roads and water bodies) from social media images, which can be translated into spatial weighting factors [11].

Motivated by this gap, we propose an inference framework that combines geographic context with multimodal social media signals to produce a landslide likelihood map constrained by toponym footprints. Our contributions are as follows.

GeoExT 2026: Fourth International Workshop on Geographic Information Extraction from Texts at ECIR 2026, April 2, 2026, Delft, The Netherlands

*Corresponding author.
bing.xu@uni-jena.de (B. Xu); alexander.brenning@uni-jena.de (A. Brenning)
0009-0001-9671-5478 (B. Xu); 0000-0001-6640-679X (A. Brenning)

- Integrating candidate toponym footprints, a landslide-susceptibility prior, and external evidence yields a spatial likelihood map that explicitly represents within-footprint location uncertainty.
- Translating CLIP-based zero-shot cues (roads and water bodies) into adaptive buffer weights enables lightweight multimodal fusion.

2. Data

We use landslide-related social media posts from X (previously Twitter) detected by the Global Landslide Detector (GLD) over the contiguous United States (CONUS) from January 1, 2020 to August 13, 2024 [1]. After manual cleaning to remove non-landslide and weak-evidence posts, we obtain 629 English-language posts with text, posting time, and attached media. For evaluation, we manually derived a reference location for each post based on the post text/media and external sources. Specifically, we inspected the described place context using Google Earth imagery and, when available, corroborated the location with online reports linked in the post or found independently. Each reference location was recorded as a point coordinate and assigned a four-level confidence label (high/medium/low/very low) following [12].

We fuse three external geographic datasets: a national-scale USGS landslide susceptibility raster for CONUS (90 m) as the spatial prior [13], PRISM daily precipitation rasters (4 km) as rainfall evidence [14], and road/water layers from OpenStreetMap plus a supplemental hydrography dataset to build proximity-based weight maps [15].

3. Method

Our goal is to produce a landslide likelihood map within the candidate toponym footprint. An overview of the framework is provided in Appendix A. All parameters are fixed across experiments and are listed in Table 2 to facilitate reproducibility. The workflow has three steps.

- Extract event-relevant place names from post text and resolve them to candidate footprints that define the search area.
- Within the search area, combine the landslide susceptibility prior with road and water-body proximity evidence. Use image cues identified by CLIP [11] to adapt the buffer weights and produce a probability surface.
- Select a reference date to retrieve a rainfall raster and resample it to the susceptibility grid. Add rainfall as a pixel-level factor to refine the spatial score, and then normalize the result to obtain the final likelihood map.

3.1. Toponym Extraction

We extract and resolve event-relevant place names using a Mistral-7B model fine-tuned with LoRA [9]. We also use a geography-guided prompting strategy to improve place-name recognition and filter for event relevance [10]. The prompt is shown in the Appendix. We then query OSM Nominatim to resolve each place name into candidate geographic objects (points, lines, or polygons). These candidates define the footprint that constrains subsequent spatial analysis and probability-map generation.

For spatial operations, we project all candidate geometries to a meter-based working coordinate system (EPSG:5070) to support buffering and area calculations. We define the analysis area based on geometry type. For point candidates, we use a fixed-radius buffer. For line candidates, we apply a line buffer. For polygon candidates, we use the polygon itself. This analysis area serves as the spatial boundary for extracting road and water features and for raster-based fusion.

3.2. Spatial probability surface

Within the analysis area, we extract road and water features from OpenStreetMap and add a supplemental hydrography dataset to improve coverage. To capture the association between landslides and nearby infrastructure and hydrologic context, we build separate buffers around roads and water bodies and rasterize them into a weight layer $W(x)$. Pixels inside a buffer receive weights greater than 1, while pixels outside receive weight 1. Where road and water buffers overlap, we take the maximum weight to avoid excessive amplification. We then clip the landslide susceptibility raster [13] to the analysis area to obtain $S(x)$, and multiply it by $W(x)$ to produce a spatial score.

3.3. CLIP-based semantic modulation

The importance of road and water cues varies across posts. We therefore use CLIP for zero-shot recognition on the attached images and obtain per-image probabilities for road and water cues, together with a landslide confidence score. If at least one image exceeds a landslide-confidence threshold, we aggregate the road and water probabilities and use them to modulate the buffer weights. For posts with multiple images, we combine a noisy-or rule with a weighted average to produce a single probability for each cue. We then map these probabilities to multiplicative gain factors applied to the road and water components $W(x)$.

3.4. Rainfall evidence and probability-map generation

To incorporate a triggering signal, we select a reference date and add the PRISM daily rainfall raster [14] from that date to strengthen spatial scoring. If the post text contains a relative time expression, such as yesterday, n days ago, or a weekday, we convert it to a calendar date relative to the posting date. Otherwise, we search a backward window of 10 days from the posting date; for each day we compute the 95th percentile of PRISM rainfall over the analysis area and select the day with the maximum value as the reference date. The rainfall factor is defined as $R(x) = 1 + \alpha \cdot \frac{\mathrm{rain}(x)}{100}$, where $\mathrm{rain}(x)$ denotes the rainfall intensity at location x, and α controls the contribution of rainfall.

The final score is defined as

$$\mathrm{Score}(x) = S(x) \cdot W(x) \cdot R(x). \tag{1}$$

4. Experiments and results

This section evaluates whether the proposed method produces uncertainty-aware landslide likelihood maps that concentrate high scores near the true location. We start from the 629 posts in Section 2 and exclude 21 posts with very low-confidence reference locations and 79 posts whose toponym resolution yields only country/state-level footprints, leaving 529 posts for evaluation. We report (i) *in-region* results for posts whose reference pixel lies within the generated map, and (ii) *overall* results that treat out-of-region cases as misses (`PR=0`, `Hit=0`), reflecting end-to-end performance.

4.1. Experimental design

We use the percentile rank (PR) of the reference pixel as the primary metric. PR is computed over finite pixels only as the empirical fraction of valid pixels whose likelihood values are $\leq$ the reference value (ties counted). To quantify search-space reduction, we report `Hit@Top`k`%` and `Area@Top`k`%`, where the top-k% region is defined by a value threshold at the $(100-k)$th percentile of valid pixel values; `Hit@Top`k`%` indicates whether the reference pixel exceeds this threshold, and `Area@Top`k`%` is the area (km^2) of pixels above the threshold. We compare three configurations:

- Susceptibility, which uses only the susceptibility raster $S(x)$.
- Rainfall, which uses only the rainfall factor $R(x)$.
- Full, which augments $S(x)$ with proximity weights for roads and water bodies, modulated by CLIP-based image cues, together with a rainfall factor $R(x)$.

Table 1
Likelihood map evaluation for Susceptibility, Rainfall, and Full (higher PR and Hit values are better; Area is in km^2).

	In-region (n = 237)			Overall (n = 441)		
Metric	Susceptibility	Rainfall	Full	Susceptibility	Rainfall	Full
PR mean	0.7154	0.6322	0.7633	0.3845	0.2454	0.4102
PR median	0.7863	0.6653	0.8581	0.2040	0.0000	0.2630
`Area@Top1%` median	15.83	7.08	2.63	–	–	–
`Area@Top5%` median	27.35	26.55	13.09	–	–	–
`Area@Top10%` median	31.76	39.69	26.16	–	–	–
`Area@Top15%` median	56.33	48.74	55.92	–	–	–
`Hit@Top1%`	18.99%	8.47%	18.14%	10.20%	3.29%	9.75%
`Hit@Top5%`	22.78%	15.68%	30.80%	12.24%	6.09%	16.55%
`Hit@Top10%`	30.80%	21.61%	42.62%	16.55%	8.39%	22.90%
`Hit@Top15%`	38.82%	27.12%	50.63%	25.85%	10.53%	31.07%

4.2. Results

Table 1 summarizes the performance of the likelihood map for the Susceptibility, Rainfall, and Full settings. For all three settings, we successfully generated maps for 441 of the 529 posts. The remaining 88 posts were excluded at this stage because the identified toponyms could not be successfully resolved to valid geographic footprints, preventing the generation of the map. Among the 441 posts with generated maps, 237 posts have a reference pixel within the generated map (in-region), while 204 reference locations fall outside the map coverage. We therefore report both in-region and overall metrics.

Overall, Full outperforms Susceptibility in both PR and hit-based metrics. Under the in-region setting, Full achieves a mean/median PR of 0.7633/0.8581, compared with 0.7154/0.7863 for Susceptibility. This indicates that, after integrating multiple evidence sources, the reference pixel is more likely to fall in a high-probability area. Gains are more pronounced for `Hit@Top5%` and `Hit@Top10%`. `Hit@Top5%` increases from 22.78% to 30.80%, and `Hit@Top10%` rises from 30.80% to 42.62%. Under the stricter overall evaluation, Full maintains consistent improvements.

By contrast, the Rainfall setting performs worse than Susceptibility across both PR and hit-based metrics, with an overall PR mean of 0.2454 and `Hit@Top10%` of 8.39%. This suggests that rainfall alone has limited spatial discriminative power and is more effective as a complementary triggering signal when fused with other evidence.

Full also substantially reduces the size of high-probability search regions. Under the in-region setting, the median `Area@Top1%` drops from 15.83 km^2 to 2.63 km^2, showing that probability mass is concentrated into a much smaller area. The median `Area@Top5%` decreases from 27.35 km^2 to 13.09 km^2, further demonstrating the advantage of Full for actionable search-space reduction.

5. Conclusion and future work

We propose a framework for generating landslide likelihood maps within toponym footprints by fusing a susceptibility prior, proximity-based evidence, semantic image cues, and rainfall signals. Experiments show that, compared with the Susceptibility setting, the Full model improves localization accuracy, with higher `Hit@Top5%` and `Hit@Top10%`, while substantially reducing the size of high-probability regions (e.g., lower `Area@Top1%`). The rainfall-only ablation indicates that rainfall alone is insufficient for precise localization but contributes when combined with other evidence. Together, these results demonstrate improved spatial focus and more actionable search-space reduction.

A key limitation is footprint mismatch, as unresolved toponyms and out-of-region cases hinder reliable end-to-end localization. Future work will focus on improving event-relevant toponym selection, fusing multiple candidate footprints, and further quantifying the contribution of each evidence source.

Declaration on Generative AI

The authors used ChatGPT to refine and polish the manuscript text. The manuscript was subsequently reviewed and revised by the authors, who take full responsibility for the final content.

References

[1] F. Ofli, U. Qazi, M. Imran, J. Roch, C. Pennington, V. Banks, R. Bossu, A real-time system for detecting landslide reports on social media using artificial intelligence, in: T. Di Noia, I.-Y. Ko, M. Schedl, C. Ardito (Eds.), Web Engineering, Springer International Publishing, Cham, 2022, pp. 49–65.

[2] M. Imran, C. Castillo, F. Diaz, S. Vieweg, Processing social media messages in mass emergency: A survey, ACM Computing Surveys 47 (2015) 67:1–67:38. doi:10.1145/2771588.

[3] F. Alam, F. Ofli, M. Imran, Crisismmd: Multimodal twitter datasets from natural disasters, in: Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018), 2018, pp. 465–473.

[4] A. S. Paradkar, C. Zhang, F. Yuan, A. Mostafavi, Examining the consistency between geo-coordinates and content-mentioned locations in tweets for disaster situational awareness: A hurricane harvey study, International Journal of Disaster Risk Reduction 73 (2022) 102878. doi:10.1016/j.ijdrr.2022.102878.

[5] A. Hürriyetoğlu, N. Oostdijk, A. van den Bosch, Estimating time to event from tweets using temporal expressions, in: Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM) @ EACL 2014, Gothenburg, Sweden, 2014, pp. 8–16.

[6] X. Hu, Z. Zhou, H. Li, Y. Hu, F. Gu, J. Kersten, H. Fan, F. Klan, Location reference recognition from texts: A survey and comparison, ACM Computing Surveys 56 (2024) 112:1–112:37. doi:10.1145/3625819.

[7] L. Richter, J. Geiß, A. Spitz, M. Gertz, Heidelplace: An extensible framework for geoparsing, in: Proceedings of the 2017 EMNLP System Demonstrations, 2017, pp. 85–90.

[8] P. Reichenbach, M. Rossi, B. D. Malamud, M. Mihir, F. Guzzetti, A review of statistically-based landslide susceptibility models, Earth-Science Reviews 180 (2018) 60–91. doi:10.1016/j.earscirev.2018.03.001.

[9] X. Hu, J. Kersten, F. Klan, S. M. Farzana, Toponym resolution leveraging lightweight and open-source large language models and geo-knowledge, International Journal of Geographical Information Science (2024) 1–28. doi:10.1080/13658816.2024.2405182.

[10] Y. Hu, G. Mai, C. Cundy, K. Choi, N. Lao, W. Liu, G. Lakhanpal, R. Z. Zhou, K. Joseph, Geo-knowledge-guided GPT models improve the extraction of location descriptions from disaster-related social media messages, International Journal of Geographical Information Science 37 (2023) 2289–2318. doi:10.1080/13658816.2023.2266495.

[11] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al., Learning transferable visual models from natural language supervision, in: International conference on machine learning, PmLR, 2021, pp. 8748–8763.

[12] L. Giglio, W. Schroeder, C. O. Justice, The collection 6 modis active fire detection algorithm and fire products, Remote Sensing of Environment 178 (2016) 31–41. doi:10.1016/j.rse.2016.02.054.

[13] G. M. Belair, J. M. Jones, S. N. Martinez, B. B. Mirus, N. J. Wood, Slope-relief threshold landslide susceptibility models for the United States and Puerto Rico, U.S. Geological Survey data release, 2024. doi:10.5066/P13KAGU3.

[14] C. Daly, M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, P. P. Pasteris, Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States, International Journal of Climatology 28 (2008) 2031–2064. doi:10.1002/joc.1688.

[15] M. Haklay, P. Weber, OpenStreetMap: User-generated street maps, IEEE Pervasive Computing 7 (2008) 12–18. doi:10.1109/MPRV.2008.80.

A. Additional Figure

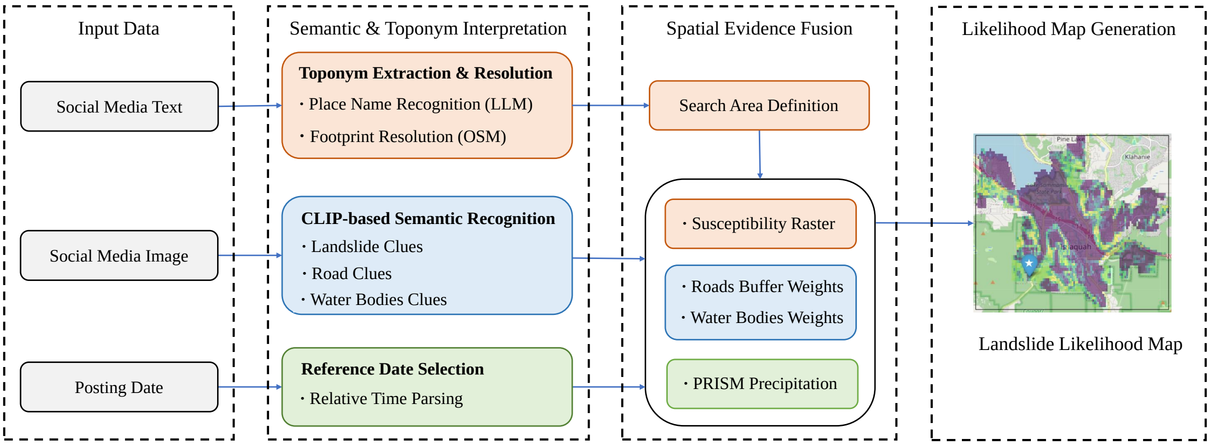

Figure 1: Overview of the proposed spatiotemporal inference framework.

B. Additional Parameters

Table 2
Key parameters used in the spatiotemporal inference framework (Full setting).

Component	Parameter	Value
Working CRS	Projection for geometry ops	EPSG:5070 (meters)
Search area	Point toponym buffer radius Line toponym buffer radius Polygon footprint	5000 m (5 km) 3000 m (3 km) 0 m
Proximity buffers	Road buffer width River buffer width	100 m 200 m
CLIP modulation	Landslide gate threshold Mapping (linear)	0.5 $m = 1 + k \cdot p$, $k_{road} = 0.6$, $k_{river} = 0.6$
Reference date selection	Lookback window Rainfall statistic	10 days 95th percentile over footprint
Rainfall fusion	Resampling (PRISM → sub-grid) Rain factor α in $R(x)$	bilinear $R(x) = 1 + \alpha \cdot \mathrm{rain}(x)/100$ 1.0

C. Additional Mistral Prompt

Instruction: This task involves recognizing **only the locations directly affected by a landslide** described in a sentence (usually a tweet + additional information). Each location should be expressed using **five hierarchical administrative fields**, even if some are unknown.

Instructions:

- Only extract **locations where the landslide occurred or caused damage**.
- For each location, return the following fields in order, separated by commas: `Street or Place, City or Town, County, State or Region, Country`
- Do not return any explanations, sentence fragments, or formatting outside the location.
- If there are multiple locations, return them as separate lines, one per location.

Example 1
Sentence: A landslide near Elkhorn City in Pike County caused the derailment.
Output: `, Elkhorn City, Pike County, Kentucky, USA`

Example 2
Sentence: Massive rainfall caused a mudslide in Southern California.
Output: `, , , California, USA`

Example 3
Sentence: A home was destroyed in Issaquah after a landslide.
Output: `, Issaquah, King County, Washington, USA`

Example 4
Sentence: Engineers are assessing a major landslide at East Ridge.
Output: `, East Ridge, Hamilton County, Tennessee, USA`

Input:
Sentence: `{text}`
A:

D. CLIP Prompt Dictionary

Prompt dictionary used for CLIP zero-shot scoring. We use the following Python dictionary to define positive/negative textual prompts for each visual cue.

```
prompt_dict = {
    "has_landslide": [
        "a photo of a landslide", "a collapsed slope", "a photo of a mudslide",
        "a photo with no landslide", "a photo of an intact hill",
        "a photo with no mudslide"
    ],
    "has_road": [
        "a photo of a road", "a highway in the mountains",
        "a photo mountains near a road", "a photo with no road",
        "a photo of nature with no road", "a photo mountains not near a road"
    ],
    "has_water": [
        "a photo of water", "a photo of a river", "a photo of a lake",
        "a photo with no water", "a photo with no river", "a photo with no lake"
    ]
}
```