IRRISIGHT: A Large-Scale Multimodal Dataset and Scalable Pipeline to Address Irrigation and Water Management in Agriculture

Nibir Chandra Mandal*
University of Virginia
wyr6fx@virginia.edu

Oishee Bintey Hoque*
University of Virginia
oishee30@virginia.edu

Mandy L Wilson
University of Virginia
alw4ey@virginia.edu

Samarth Swarup
University of Virginia
swarup@virginia.edu

Sayjro K Nouwakpo
Agricultural Research Service
United States Department of Agriculture
kossi.nouwakpo@usda.gov

Abhijin Adiga
University of Virginia
abhijin@virginia.edu

Madhav Marathe
University of Virginia
marathe@virginia.edu

Abstract

The lack of fine-grained, large-scale datasets on water availability presents a critical barrier to applying machine learning (ML) for agricultural water management. Since there are multiple natural and anthropogenic factors that influence water availability, incorporating diverse multimodal features can significantly improve modeling performance. However, integrating such heterogeneous data is challenging due to spatial misalignments, inconsistent formats, semantic label ambiguities, and class imbalances. To address these challenges, we introduce IRRISIGHT, a large-scale, multimodal dataset spanning 20 U.S. states. It consists of 1.4 million pixel-aligned 224×224 patches that fuse satellite imagery with rich environmental attributes. We develop a robust geospatial fusion pipeline that aligns raster, vector, and point-based data on a unified 10m grid, and employ domain-informed structured prompts to convert tabular attributes into natural language. With irrigation type classification as a representative problem, the dataset is AI-ready, offering a spatially disjoint train/test split and extensive benchmarking with both vision and vision–language models. Our results demonstrate that multimodal representations substantially improve model performance, establishing a foundation for future research on water availability. https://github.com/Nibir088/IRRISIGHT https://huggingface.co/datasets/OBH30/IRRISIGHT

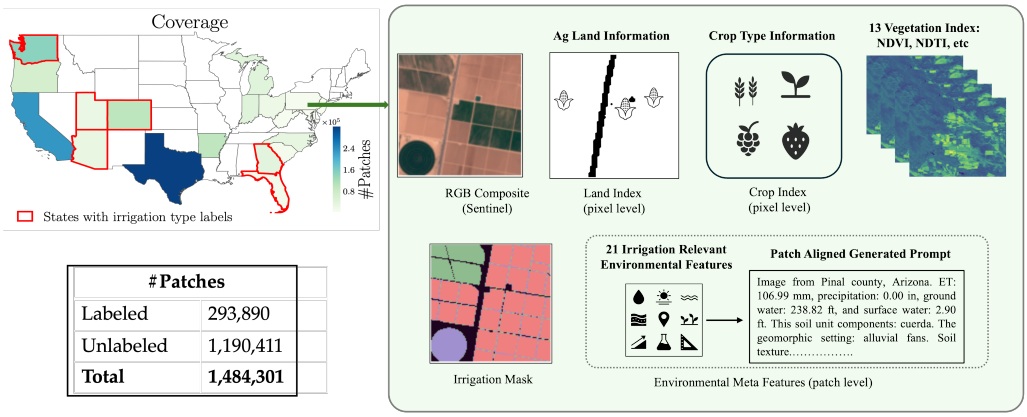
1 Introduction

Managing agricultural water use under increasing climate pressures is a growing challenge. Excessive irrigation has contributed to declining groundwater levels and reduced river discharge, posing serious threats to water security and ecosystem health [15, 79, 62, 24, 55]. A comprehensive view that integrates irrigation practices, hydrology, soil properties, and crop types is essential for understanding and managing water availability. While efficient irrigation methods can reduce water losses from

*Equal contribution.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Track on Datasets and Benchmarks.

https://doi.org/10.52202/085713-0186



#Patches	
Labeled	293,890
Unlabeled	1,190,411
Total	**1,484,301**

Figure 1: An overview of coverage and dataset structure. (a) Shows coverage of our dataset. (b) Sample multi-modal (containing 37 meta features) ML-ready data from our dataset.

runoff—and help sustain crops during drought—they often require substantial investments from farmers [42]. Mapping water availability at high spatial resolution and across large regions is therefore critical. Such maps not only inform sustainable water management but also support climate adaptation strategies, track irrigation infrastructure investment trends, and enable regional or basin-scale modeling of the environmental impacts of shifting irrigation practices [36, 54, 3].

Remote sensing, combined with deep learning, has been extensively applied for large-scale mapping of agricultural features [35, 76, 46, 45, 12, 29, 1, 32, 23, 7, 60]. From the perspective of water availability, numerous studies have focused on distinguishing irrigated from non-irrigated lands in mixed-use agricultural areas. This has led to the creation of several prominent datasets at varying spatiotemporal scales [13, 66, 63, 40, 65, 78, 33, 11]. Another important related area is cropland monitoring [32, 31], as crop type directly influences irrigation decisions—such as the choice between sprinkler, drip, or flood systems—and therefore impacts overall water consumption. Additionally, water stress monitoring during droughts has been explored using remote sensing and deep learning approaches [75]. This work builds on our previous efforts focused on irrigation type mapping [25, 27], addressing several technical challenges identified in earlier research. A more detailed comparison with these prior studies is provided in the related works section.

The lack of fine-grained, large-scale data presents a major challenge to deploying machine learning (ML) tools for optimizing agricultural water use. Although ML has achieved success in land cover and crop classification using satellite imagery, applying these methods to water availability problems is fundamentally more complex [27]. This complexity arises from three key challenges. First, water availability is influenced by irrigation decisions as well as physical factors that depend not only on visible surface features, but also on invisible factors, including soil properties, hydrological characteristics, topography, water availability, climate patterns, and crop requirements. Second, integrating these variables from heterogeneous data sources is non-trivial due to inconsistent formats, spatial misalignment, missing data patterns, and varying temporal granularity. Aligning these onto common spatial units requires reprojection, resampling, and interpolation strategies that must preserve semantic integrity while avoiding aliasing or feature leakage. The lack of standard tools and benchmarks for this fusion process increases the risk of bias and inconsistency across regions. Third, the semantic ambiguity of class labels (in the case of irrigation type mapping for e.g., drip vs. furrow vs. sprinkler) and high class imbalances in real-world settings make classification unstable and unreliable without careful curation and domain knowledge.

Contributions. We present IRRISIGHT, the first large-scale, multimodal dataset designed to enable generalizable machine learning for addressing water availability problems across the United States (US). An ML-ready sample from this dataset is shown in Figure 1. Our contributions address critical gaps in data availability, representation, and evaluation by integrating diverse datasets. This is followed by extensive benchmarking using various state-of-the-art models. A summary of our contributions is as follows.

- **Novel Large-scale Multimodal Dataset:** We construct a dataset of over 1.4 million pixel-aligned 224×224 geospatial patches across twenty agriculturally important US states. Each patch integrates visual and non-visual modalities, including soil surveys, climate variables (precipitation, evapotranspiration (ET)), and hydrologic data (groundwater, surface water). To our knowledge, this is the first dataset to align such diverse modalities for water availability mapping.
- **Automated Data Fusion and Standardization Pipeline:** We develop a fully automated pipeline (see Figure 2) that integrates heterogeneous geospatial inputs—raster, vector, and point-based data—onto a common 10m grid using spatial joins, reprojection, masking, and resolution-aware resampling. Our pipeline processes 6.8TB of raw geospatial data from 11 public sources, performing standard filtering, spatial cropping, and quality control across the twenty states across the contiguous US (covering 155 million acres of agricultural land). This scalable pipeline allows us to continuously expand our geographic coverage.
- **Text-augmented Representations via Structured Prompts:** We convert non-visual features (e.g., soil texture, drainage, slope, hydrologic group) into natural language descriptions using a domain-informed, rule-based system. These prompts enable alignment with vision–language models and inject agronomic context that is not visible in imagery. Unlike one-hot or numerical feature vectors, natural language provides a semantically dense, interpretable format that supports alignment with multimodal models, facilitates human-in-the-loop analysis, and allows flexible conditioning in generative or contrastive learning settings. Thus, the text modality serves not just as auxiliary metadata, but as a linguistically structured, domain-informed signal that bridges spatial data with real-world irrigation knowledge.
- **ML-Ready Dataset and Evaluation Framework:** Using irrigation type detection as a representative problem, our work uses labeled data from six states to develop a training dataset. We provide a spatially disjoint training/testing split dataset, and an evaluation setup for states with irrigation type labels. After processing and alignment, we have a total of **1,484,301 ML-ready patches** across 20 states, with **293,890 labeled** and **1,190,206 unlabeled** examples. Spatial correlations and out-of-distribution generalizations are accounted for in the train-validate-test splitting. We also provide model-predicted labels with corresponding confidence scores for unlabeled data.
- **Extensive Benchmarking with Vision and Vision–Language Models:** We benchmark a range of models, including convolutional neural networks (CNNs), transformers, and vision-language models on our dataset for cross-state supervised learning. Our experiments demonstrate that the multimodal inputs with structured text prompts significantly improve cross-state generalization compared to image-only baselines on labeled data.

2 Related Work

Existing irrigation mapping products derived by remote sensing. A range of remote sensing-based datasets have been created to distinguish irrigated from non-irrigated areas without providing information on specific irrigation methods. Recent efforts to map irrigation have increasingly focused on higher-resolution datasets. For instance, the AIM-HPA dataset provides 30-meter resolution irrigation maps, though it is limited to the High Plains region [11]. LANID [78] applies decision-tree methods to generate annual irrigation maps over a 20-year span, accompanied by open-access ground-truth data. IrrMapper [33] further leverages 60,000 point samples collected over 28 years, integrating Landsat imagery with climate, meteorological, and terrain features to train a Random Forest classifier. At coarser resolutions, the Moderate Resolution Irrigated Area Dataset (MIrAD) offers 250-meter resolution maps but relies heavily on census-based statistics, which can lead to considerable classification errors [6, 56]. However, none of these works address the problem of identifying irrigation types, nor do they apply state-of-the-art ML techniques.

Related work from our team. This work is part of a series of works undertaken by our team in mapping irrigation infrastructure [27, 25, 26, 43]. Our recent work [27] – Knowledge-Informed Irrigation Mapping (KIIM) – takes the first steps toward integrating multimodal data for irrigation classification. It contributes methods to improve classification performance with limited labeled data and poor spatial coverage. Its precursor [25] – IrrNet – is a preliminary work undertaken where only satellite imagery is used for the purpose. Our current work is motivated by the lessons learned and challenges faced in these works. KIIM focuses on states with labeled data, and, as a result, is limited in terms of scalability due to several structural constraints. It incorporates state-specific priors using a

 https://doi.org/10.52202/085713-0186

projection matrix, which is unavailable for most of the states. We consolidated the dataset used in KIIM and released it as IrrMap [43]. IrrMap includes labeled data for four states, along with satellite image patches, crop masks, derived indices, and land-use information. IRRISIGHT improves upon IrrMap in both geographic coverage and feature richness. It incorporates labeled data from six states and augments it with additional textual information from soil databases and hydrological sources using LLMs. This multimodal dataset is also used to benchmark a broader set of models—including visual–LLM models—beyond those evaluated in IrrMap. While IRRISIGHT builds on the foundation of IrrMap, which focused specifically on irrigation type detection, it is designed to support a wider range of water availability and management questions. In summary, while prior works such as IrrMap and KIIM laid important groundwork in irrigation mapping and modality fusion, IRRISIGHT advances the field by providing a large-scale, multimodal, and extensible benchmark designed for modern vision–language and geospatial learning (see Section C in supplement).

Works on remote-sensing and multimodal data. Recent work in multimodal remote sensing has introduced large-scale foundation models trained on diverse data types such as optical, SAR, and multispectral imagery to support universal earth observation tasks [22, 16] and vision-language applications like captioning [38], question answering [34, 30], and region-level reasoning [51]. Generative and contrastive frameworks further enable effective data assimilation and cross-modal representation learning [58, 39]. These advancements are complemented by emerging benchmark datasets that span modalities, spatial resolutions, and temporal dynamics, facilitating research in tasks such as segmentation [50, 18], time-series forecasting [4, 5] and multimodal text-to-image generation [41, 19]. In agriculture, these multimodal approaches improve monitoring by combining optical and multispectral data to capture crop and land use dynamics [9, 10, 47, 64].

3 Data Collection

Our dataset integrates geospatial data from multiple public sources, spanning raster, vector, and point-based formats. All data are harmonized to the Albers Equal Area Conic projection (EPSG:5070) for spatial consistency. Table 1 summarizes the core data sources used. Below, we provide some details about the data and its processing. More details are provided in the supplement.

Satellite imagery: Sentinel-2 imagery was obtained from USGS Earth Explorer. Data acquisition focused on the peak irrigation season (July), a key period for assessing water use and crop conditions, as imagery from other periods may capture snow cover, bare soil, or dormant vegetation, making it less useful for irrigation analysis. Images were filtered out if they did not meet quality criteria or were deemed redundant.

Irrigation labels: No national-scale irrigation label dataset is available. This data was obtained for each of the six states for which it is publicly available (See Table in Supplement). These datasets vary significantly in terms of temporal coverage, spatial granularity, and annotation standards. While all sources include field-level polygons and associated attributes, such as irrigation status, method type, crop classification, and water source, differences in data collection protocols introduce substantial label noise. For Utah, Washington, and Colorado, we have complete statewide coverage of irrigated lands. Furthermore, irrigation practices vary considerably by state. Notable class imbalances across and within states pose additional challenges for model training.

Land use and crop type: We utilize two national-scale raster products to identify and contextualize agricultural areas. The MRLC National Land Cover Database (NLCD) offers 30m-resolution land cover classifications across 16 categories, including cultivated cropland and pasture (See Table 1). Complementarily, the USDA Crop Data Layer (CDL) provides annual, per-pixel crop type labels from 2008 to 2023, with over 130 unique crop classes. Together, these layers are used to filter cropland regions and enrich patch-level annotations with crop-specific information.

Water availability: We incorporate environmental observations from the USGS National Water Information System (NWIS), which provides raw, site-specific measurements of groundwater depth, surface water elevation, and precipitation at daily temporal resolution. We preprocess this data by filtering for relevant sites within our study regions and aggregating the daily values into monthly averages spanning 2010 to 2025.

Evapotranspiration: We incorporate monthly evapotranspiration (ET) rasters from the USGS Famine Early Warning Systems Network (FEWS). This data provides regional-scale estimates of

water loss through soil evaporation and plant transpiration at 1km resolution. While ET is a critical factor for understanding irrigation needs, it is challenging to integrate with fine-resolution satellite imagery due to its coarse spatial granularity and atmospheric dependencies. Nonetheless, we spatially align ET values to image patches to provide a proxy for water stress and latent demand during peak growing seasons.

Soil data: We utilize detailed soil information from the USDA NRCS SSURGO database. The database contains a nationwide geospatial dataset compiled through extensive field surveys, lab analysis, and expert interpretation. The data are provided as vector polygons (map units), where each map unit is linked to a set of relational tables describing soil components, horizon-level measurements, texture groups, and geomorphic features. We use attributes such as slope, hydrologic group, composition percentage, organic matter, bulk density, soil texture descriptions, and geomorphic landform classifications. These properties are crucial for assessing soil water retention, permeability, and suitability for different irrigation methods.

Table 1: Summary of Data Sources Used in Dataset Construction.

Source	Description	Spatial Resolution	Temporal Coverage	Use
Sentinel-2 [2]	Multispectral satellite imagery (10 bands)	10m (20m bands upsampled)	2014–2023 (July only)	Visual input for patch extraction
USGS Irrigated Lands [69]	State-level irrigation maps	30m raster	2002–2017	Supervised labels for irrigation type
MRLC NLCD [48]	Land Use / Land Cover (LULC) classification	30m raster	2014–2023	Filtering non-agricultural patches
USDA CDL [71]	Crop Data Layer with per-pixel crop types	30m raster	2008–2023	Crop-based filtering, auxiliary input for prompts
USGS NWIS [72]	Groundwater, surface water, and precipitation (station-level)	Point data (lat/lon)	2010–2025 (monthly)	Patch-level climate/hydrology features
USGS FEWS ET [73]	Monthly evapotranspiration rasters	1km raster	2014–2023	ET features per patch
USDA NRCS SSURGO [52]	Soil map units, components, horizons, texture, geomorphology	1:24,000 scale vector polygons	Static	Text prompt generation, irrigation suitability assessment
US Census TIGER/Line [70]	County boundary polygons with metadata (name, state, FIPS)	~1:500,000 scale vector polygons	Static	Regional referencing, county-level context encoding

4 Data Processing Pipeline

An outline of the various steps involved in data acquisition, processing, and integration is provided in Figure 2. Similarly, an outline of the preparation of the ML-ready dataset is provided in the supplement. We provide a summary of each data processing step below (detailed descriptions are provided in the supplement).

4.1 Data Acquisition

All datasets are summarized in Table 1. We retrieve Sentinel-2 metadata from the Copernicus Data Space Catalogue, filtering for cloud-free acquisitions within the growing season window. We remove duplicate scenes to ensure a balanced spatial coverage across the dataset. The associated satellite imagery for the selected scenes are then downloaded via authenticated API requests. Using state-level polygons, we extract overlapping land cover (from NLCD), crop type (from CDL), and soil information (SSURGO). Extracted land use and crop type rasters are reprojected to a 10m grid via nearest-neighbor interpolation. The SSURGO soil data provides polygonal soil map units with associated horizon-level and component-level attributes. For hydrological data, we queried the USGS National Water Information System to collect data for all active monitoring sites across the US states, including geolocation and site identifiers.

4.2 Data Processing

Satellite imagery and other land masks. Sentinel-2 scenes are parsed into ten multispectral bands (B02–B12) spanning 10m and 20m resolutions. These are aligned using equal-area projection and bilinear sampling. Then the stacked bands are divided into fixed-size tiles of 224×224 pixels. We retain patches only if at least 10% of pixels fall into cropland categories and contain valid reflectance

 https://doi.org/10.52202/085713-0186

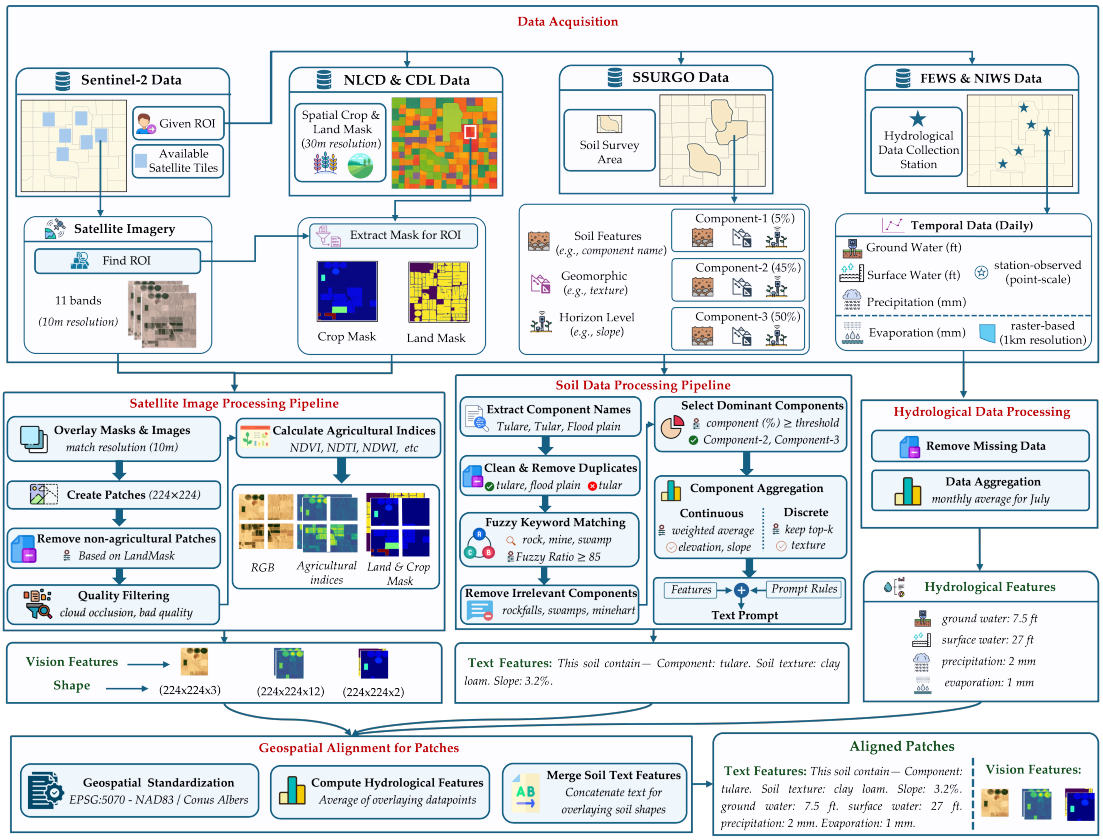


Figure 2: Multimodal Data Processing Pipeline.

values. All land use and crop masks are reprojected using nearest-neighbor interpolation to preserve categorical semantics.

Derived indices. To enhance surface property analysis for irrigation mapping, we compute a suite of spectral indices capturing vegetation health, water presence, and soil conditions. These include popular indices such as NDVI, GNDVI, CIgreen, NDWI, EVI, SAVI, and MSAVI. We have provided the details in the supplement.

Soil data. We standardize gSSURGO soil survey data to construct spatially aligned soil features suitable for ML tasks. Each mapped unit contains multiple overlapping soil components, each with its own attributes, including: (i) *horizon-level measurements* depth-specific properties such as available water capacity, organic matter, and hydraulic conductivity; (ii) *soil texture*: the relative proportions of sand, silt, and clay; and (iii) *geomorphic attributes* landscape positions such as terraces or floodplains. Each component differs in spatial dominance, which is defined by its areal proportion within the map unit. To ensure consistency, we follow a structured aggregation technique by selecting irrigation-relevant components with significant spatial coverage. The details of filtering and aggregation are in the supplement.

Hydrological data. For each hydrological site, we retrieved daily measurements of gauge height, groundwater level, or precipitation, and aggregated them (by taking the mean) into monthly means from 2016 to 2025. Each time series was aligned to a complete (2016–2023) year–month index to preserve temporal consistency and explicitly represent missing data.

4.2.1 Text-Prompt Generation

To supplement satellite imagery with non-visible but agronomically critical information not directly observable at the pixel level, we generate structured, rule-based natural language prompts that encode

localized soil and landform characteristics. These prompts enrich the feature space with domain knowledge reflecting soil and landscape attributes that influence irrigation suitability.

Each image patch is spatially joined with soil map units from the USDA NRCS **SSURGO** database. These polygons include: (i) a list of soil components (e.g., *Hanford loamy sand*) with their areal composition percentages, and (ii) attributes for each component, such as texture, drainage class, hydrologic group, slope range, geomorphic description, and horizon-level properties (e.g., bulk density, available water capacity).

For each patch, we select up to two dominant components per map unit based on areal proportion (typically $>30\%$), representing the most influential soil types within that region. Structured attributes are then converted into coherent text using handcrafted templates. For example, attributes such as soil name (*Tulare*), drainage class (*well drained*), texture group (*sandy loam*), runoff class (*moderate*), and slope (*averages 3.2%*) are concatenated into a standardized template:

> "This soil unit contains Tulare. Soil texture includes sandy loam. It is classified as well drained with moderate runoff. The average slope is 3.2%."

If multiple soil components intersect a patch, their corresponding prompts are concatenated using a delimiter (`##`) to preserve provenance and reduce ambiguity. The resulting textual prompt is assigned to each image patch as a field named `text_prompt`.

4.3 Data Integration

To construct a unified multimodal dataset, we spatially align all heterogeneous geospatial inputs—satellite imagery, environmental variables, and soil attributes—into structured 224×224 patches. Each patch combines pixel-level reflectance with auxiliary domain features, including evapotranspiration (ET), precipitation, groundwater depth, surface water elevation, and processed soil properties. All inputs are reprojected to a common projection EPSG:5070. Patch-level values are extracted using: (i) polygonal intersection for irrigation labels and soil map units; (ii) centroid-based nearest neighbor lookup for point data (e.g., groundwater, precipitation); and (iii) raster sampling for gridded layers (e.g., ET). Soil prompts are joined to each patch via spatial intersection; if multiple soil units overlap, prompts are concatenated.

4.4 ML-ready Dataset Preparation

Here, we consider the representative problem of irrigation type mapping. However, our data can also be applied to related tasks by reusing the same processing pipeline provided with the dataset. After processing and alignment, we extract a total of **1,484,301 ML-ready patches** across 20 states, with **293,890 labeled** and **1,190,206 unlabeled** examples. The labeled subset comes from six states, some with partial coverage (See Figure 1), enabling semi-supervised and pretraining applications. After reprojecting, masking, and patch extraction, we obtain **5,649 processed tiles** occupying **1,386.5 GB**, which represents a **77% reduction from raw data** in storage size compared to the original data footprint.

Training and evaluation splits. We adopt two complementary strategies for splitting our dataset to support both intra-state generalization and cross-regional evaluation: a 70:15:15 spatial split within states, and a leave-one-state-out protocol across states. In the cross-state splitting, for each labeled state the train-validate-test split is performed at the tile level (one satellite image) as opposed to splitting the set of individual patches. It ensures that data from the same geographic region does not appear in multiple splits—thereby preventing spatial leakage. This design is critical in remote sensing settings, where nearby patches are highly correlated due to shared land cover and climate conditions. To evaluate out-of-distribution generalization, we perform a leave-one-state-out split. A target state is held out entirely for testing, while the remaining states contribute data to the training and validation sets. For the training states, we again split at the tile level and reserve a small portion (typically 10%) for validation. This setting simulates real-world scenarios where labeled data is unavailable in a new geographic region, and the model must transfer knowledge learned from other states. The details are in the supplement.

https://doi.org/10.52202/085713-0186

5 Dataset Benchmarking

We evaluate segmentation performance across three irrigation types (Flood, Sprinkler, and Drip) using nine baseline models and multiple modalities using standard Dice and IoU metrics (details in the supplement). Table 2 reports overall metrics, while Table 3 details per-state Dice scores under the leave-one-state-out setting.

Architecture Comparison. Among RGB-only architectures, transformer-based models (e.g., SegFormer, FarSeg) outperform CNN baselines. However, performance improvements from architectural changes are modest compared to those achieved by integrating additional modalities. Vision-language models like RemoteCLIP leverage text supervision to improve generalization, while KIIM benefits from structured auxiliary inputs (crop and land use). KIIM (RGB + Crop + Land) achieves the highest performance across all types, with Dice scores of 93.6% (Flood), 95.8% (Sprinkler), and 94.6% (Drip). RemoteCLIP and CLIP, which incorporate textual prompts, also surpass all RGB-only models. This indicates the utility of language-grounded supervision. These results underscore that incorporating domain-specific context is more impactful than architecture alone for the irrigation mapping task.

Cross-State Generalization. KIIM maintains strong performance across diverse states, likely due to its explicit fusion of structured crop and land metadata with visual features, while RGB-only models degrade significantly in out-of-distribution regions. CLIP and RemoteCLIP perform well in states with clearer visual-text alignment (e.g., AZ, CO) but struggle in regions with less distinctive irrigation patterns (e.g., WA, FL), which highlights the limits of vision-only and VLM-based generalization without structured context. Among the irrigation types, sprinkler irrigation is consistently the easiest to segment (due to its circular geometric shape), with most models achieving Dice greater than 85%. Flood irrigation shows variability across states. Drip irrigation is the most challenging, particularly in CO and UT, where even strong models like KIIM score below 15%, likely due to its irregular shapes and fine-grained, low-contrast footprint.

Table 2: Performance (%) across irrigation types using different models and modalities.

Model	Modality	Flood		Sprinkler		Drip	
		Dice	IoU	Dice	IoU	Dice	IoU
ResNet	RGB	35.2	21.4	92.2	85.5	88.5	79.4
ViT	RGB	82.9	70.9	89.9	81.7	84.1	72.6
FPN	RGB	86.2	75.7	91.4	84.1	86.4	76.0
SegFormer	RGB	86.2	75.8	91.7	84.7	85.9	75.3
DeepLabV3+	RGB	87.2	77.4	92.1	85.4	86.9	76.8
FarSeg	RGB	88.2	78.9	92.3	85.7	89.2	80.5
CLIP	RGB+Text	90.1	82.0	93.1	87.0	90.7	83.0
RemoteCLIP	RGB+Text	90.9	83.3	93.7	88.2	92.3	85.7
KIIM	RGB+Land+Crop	93.6	88.0	95.8	91.9	94.6	89.7

Impact of Textual Prompts. As shown in Table 2, incorporating these text prompts substantially improves model performance compared to RGB-only baselines. In RemoteCLIP, the textual prompt serves as an independent semantic input encoded via a language encoder, complementing the visual encoder rather than being concatenated as auxiliary features. For instance, RemoteCLIP achieves a 92.3% Dice score on drip irrigation, compared to 88.5% for an RGB-only ResNet, confirming that structured domain knowledge enhances model discrimination across all irrigation types.

Label Generation for Unlabeled States. Following the performance of KIIM model, we generate synthetic labels for 17 unlabeled states. We trained the KIIM model on all six labeled data and generated irrigation maps for the unlabeled states. We show the confidence for synthetic labels in Figure 3 for seven states.

6 Application of the Dataset

The IRRISIGHT dataset is designed to support a range of machine learning and agricultural modeling tasks. A direct application is that it lays a pipeline for ingesting feature-rich ML-ready multimodal dataset that can help address at the national-scale various questions in the context of water availability.

Table 3: Dice scores (%) for Flood, Sprinkler, and Drip irrigation types across 6 U.S. states and 9 model architectures. Flood is not applicable in Georgia.

State	Type	KIIM	R-CLIP	ViT	FarSeg	CLIP	DLV3+	FPN	SegFormer	ResNet
AZ	Flood	**74.6**	67.0	72.1	70.6	66.7	70.69	71.33	70.07	70.88
	Sprinkler	83.2	82.8	**84.5**	82.8	79.7	83.01	82.43	85.62	83.31
	Drip	22.5	5.8	10.3	**29.4**	1.8	18.11	21.92	23.04	17.54
CO	Flood	52.8	12.1	**53.1**	52.2	44.1	50.31	49.44	46.98	50.52
	Sprinkler	**79.0**	73.4	76.1	78.6	73.5	76.53	77.84	78.27	76.29
	Drip	**1.9**	0.6	0.5	1.1	0.1	0.76	0.67	1.27	0.19
UT	Flood	**61.6**	28.1	60.6	60.3	54.9	59.60	57.98	60.09	59.71
	Sprinkler	**67.9**	56.4	61.5	65.3	57.5	62.97	62.59	63.68	63.55
	Drip	**14.8**	0.0	6.4	0.8	0.0	9.73	4.23	4.12	6.46
WA	Flood	**29.5**	19.7	21.6	25.6	18.3	19.91	22.83	23.65	22.23
	Sprinkler	**79.7**	74.3	77.4	78.6	73.3	76.66	77.38	77.74	77.09
	Drip	8.9	1.6	**22.1**	12.7	4.5	18.25	10.05	27.18	0.01
FL	Flood	11.6	3.0	0.8	17.3	2.2	1.60	8.46	5.11	**17.91**
	Sprinkler	**64.8**	51.0	54.5	61.0	50.6	59.76	59.58	57.57	61.14
	Drip	15.2	2.7	**30.3**	28.2	8.5	6.14	11.51	8.08	3.70
GE	Sprinkler	**59.0**	42.4	48.9	50.4	37.2	27.12	47.33	45.93	47.38
	Drip	6.2	1.6	5.5	**6.8**	3.3	1.55	6.18	11.34	2.95

While irrigation type segmentation is the primary benchmark, *IRRISIGHT* was designed for broader agricultural water management tasks. Its modular multimodal pipeline—integrating Sentinel-2 imagery with soil, hydrological, and crop metadata—enables generalization beyond irrigation mapping. To illustrate this extensibility, we evaluated two downstream tasks: *crop classification* and *environmental variable regression*. The KIIM model retrained on IRRISIGHT data significantly outperformed the USDA CropScape [49], achieving higher Macro-F1 scores across all states (Supplement Table S10). Regression experiments using tree-based models also showed strong predictive accuracy for evapotranspiration, groundwater, precipitation, and surface water (Table 4), demonstrating IRRISIGHT's utility across diverse hydrological prediction tasks.

Furthermore, regression experiments using Random Forest, Gradient Boosting, and XGBoost models were conducted to predict key environmental variables—evapotranspiration (ET), groundwater, precipitation, and surface water—using other available features such as irrigation labels and geospatial attributes.. All models achieved strong coefficients of determination (R^2 ranging from 0.43 to 0.99) across variables (Table 4), confirming that the dataset's integrated structure enables robust learning for hydrologically relevant targets.

Table 4: Regression performance (MAE, RMSE, and R^2) for environmental variables using tree-based models. Lower MAE/RMSE and higher R^2 are better.

Variable (Unit)	Model	MAE	RMSE	$\mathbf{R^2}$
ET (mm)	Random Forest	22.993	30.279	0.483
	Gradient Boosting	24.751	31.705	0.433
	XGBoost	23.075	30.328	0.481
Ground Water (ft)	Random Forest	3.230	19.544	0.867
	Gradient Boosting	16.314	36.546	0.536
	XGBoost	10.683	28.671	0.715
Precipitation (in)	Random Forest	0.001	0.005	0.988
	Gradient Boosting	0.009	0.016	0.869
	XGBoost	0.003	0.008	0.963
Surface Water (ft)	Random Forest	5.782	64.259	0.986
	Gradient Boosting	60.448	174.265	0.901
	XGBoost	27.827	125.645	0.948

These findings provide strong empirical evidence that *IRRISIGHT* generalizes beyond irrigation classification, supporting diverse supervised and regression-based tasks in agricultural water manage-

https://doi.org/10.52202/085713-0186

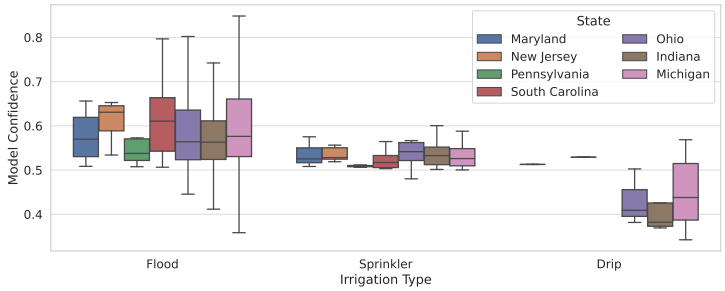


Figure 3: Distribution of model confidence scores across irrigation types (excluding non-irrigated) for the top 7 high-confidence states. Each boxplot shows the variation in predicted confidence across patches within a given irrigation class and state. Confidence values are derived from the model's softmax outputs, and higher scores indicate greater certainty in the predicted irrigation type.

ment. By design, it enables research in drought assessment, fallowing detection, irrigation suitability mapping, and crop–irrigation optimization. Some other example applications are provided below.

Irrigation Classification with Supervised and Semi-Supervised Learning: The labeled subset enables training of supervised deep learning models for irrigation detection, while the large unlabeled regions support semi-supervised approaches that leverage weak labels and spatial context.

Flash Drought Detection: A flash drought is a sudden onset drought that lasts long enough to impact vegetation [37]. This dataset can be combined with flash drought data [e.g., 53] to develop models of flash drought impact on crops and which irrigation practices are best at minimizing it.

Fallowing Detection and Prediction: Fallowing, where a field is not planted for a season, can be done as part of a normal crop rotation sequence, or due to lack of water availability (drought), or as part of a strategy for long-term water security [59, 57]. The dataset can be used to develop fallowing detection and prediction models, which can in turn be used to discover normal crop rotation patterns and to quantify the impact of drought.

Irrigation Suitability Mapping: By combining ET, precipitation, and groundwater data, the dataset can be used to estimate irrigation feasibility in areas with limited infrastructure or policy-relevant water stress.

Region-Specific Analysis for Crop Planning: Crop type labels, where available, enable downstream tasks such as identifying crop-irrigation suitability patterns and generating region-specific irrigation recommendations based on soil and water constraints.

7 Limitations

Here, we acknowledge some of the shortcomings of the IRRISIGHT datasets and describe possible future directions. A major limitation is that pixel-level labels are included for only six states. Further, not all states have full coverage. While KIIM [27] shows superior performance in irrigation type mapping, the poor results on cross-state generalization suggest the importance of providing the model with features and if possible sample labels corresponding to the target region for training to improve performance. Also, for states where limited or no label information is present, validation becomes a challenge. Using data such as US Agricultural Census, an aggregate-level validation based on county data is a possibility but has its own challenges. Raster reprojection to a 10m grid (e.g., crop type, land use) may cause misalignment or information loss, especially in heterogeneous landscapes. Our soil aggregation emphasizes dominant components, potentially removing hydrologically relevant minority types. In the future, we will explore rigorous evaluation strategies for unlabeled states, including census-model alignment and expert review.

Acknowledgments

This material is based upon work supported by the AI Research Institutes program supported by NSF and USDA-NIFA under the AI Institute: Agricultural AI for Transforming Workforce and Decision Support (AgAID) award No. 2021-67021-35344. This work was partially supported by University of Virginia Strategic Investment Fund award number SIF160.

References

[1] Aniruddha Adiga, Surbhi Singh, Ethan Choo, Johnny Yang, Srinivasan Venkatramanan, Anjana Devkota, Bharat Babu Shrestha, Seerjana Maharjan, Sita Gyawali, Sandeep Dhakal, Krishna Poudel, Pramod Kumar Jha, Rangaswamy Muniappan, Madhav Marathe, and Abhijin Adiga. A robust deep learning framework reveals the spread of multiple invasive plants in a biodiversity hotspot using satellite imagery. In *The Workshop on Artificial Intelligence for Social Good (in AAAI)*, 2023.

[2] European Space Agency. Sentinel-2: Earth observation mission. *Copernicus Programme*, 2023. URL `https://en.wikipedia.org/wiki/Sentinel-2`.

[3] US Environmenta Protection Agency. Climate change impacts on agriculture and food supply. `https://www.epa.gov/climateimpacts/climate-change-impacts-agriculture-and-food-supply`, 2020. Accessed: 2024-03-24.

[4] Matthew Allen, Francisco Dorr, Joseph Alejandro Gallego Mejia, Laura Martínez-Ferrer, Anna Jungbluth, Freddie Kalaitzis, and Raúl Ramos-Pollán. M3leo: A multi-modal, multi-label earth observation dataset integrating interferometric sar and multispectral data. *Advances in Neural Information Processing Systems*, 37:104694–104723, 2024.

[5] Vitus Benson, Claire Robin, Christian Requena-Mesa, Lazaro Alonso, Nuno Carvalhais, José Cortés, Zhihan Gao, Nora Linscheid, Mélanie Weynants, and Markus Reichstein. Multi-modal learning for geospatial vegetation forecasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27788–27799, 2024.

[6] Jesslyn F Brown and Md Shahriar Pervez. Merging remote sensing data and national agricultural statistics to model change in irrigated agriculture. *Agricultural Systems*, 127:28–40, 2014.

[7] Inacio T Bueno, João FG Antunes, Aliny A Dos Reis, João PS Werner, Ana PSGDD Toro, Gleyce KDA Figueiredo, Júlio CDM Esquerdo, Rubens AC Lamparelli, Alexandre C Coutinho, and Paulo SG Magalhães. Mapping integrated crop-livestock systems in brazil with planetscope time series and deep learning. *Remote Sensing of Environment*, 299:113886, 2023.

[8] CO-DWR. Colorado decision support system (cdss). `https://dwr.colorado.gov/services/data-information/gis`, 2025. [Accessed 03-2025].

[9] Yezhen Cong, Samar Khanna, Chenlin Meng, Patrick Liu, Erik Rozi, Yutong He, Marshall Burke, David Lobell, and Stefano Ermon. Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery. *Advances in Neural Information Processing Systems*, 35: 197–211, 2022.

[10] MRSB DATA. Multimodal artificial intelligence foundation models: Unleashing the power of remote sensing big data in earth observation. *Innovation*, 2(1):100055, 2024.

[11] Jillian M Deines, Anthony D Kendall, Morgan A Crowley, Jeremy Rapp, Jeffrey A Cardille, and David W Hyndman. Mapping three decades of annual irrigation across the us high plains aquifer using landsat and google earth engine. *Remote Sensing of Environment*, 233:111400, 2019.

[12] Maurilio Di Cicco, Ciro Potena, Giorgio Grisetti, and Alberto Pretto. Automatic model based dataset generation for fast and accurate crop and weeds detection. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5188–5195. IEEE, 2017.

 https://doi.org/10.52202/085713-0186

[13] Petra Döll and Stefan Siebert. A digital global map of irrigated areas. *ICID J*, 49(2):55–66, 2000.

[14] Rachel R Fern, Elliott A Foxley, Andrea Bruno, and Michael L Morrison. Suitability of ndvi and osavi as estimators of green biomass and coverage in a semi-arid rangeland. *Ecological Indicators*, 94:16–21, 2018.

[15] Jonathan A Foley, Navin Ramankutty, Kate A Brauman, Emily S Cassidy, James S Gerber, Matt Johnston, Nathaniel D Mueller, Christine O'Connell, Deepak K Ray, Paul C West, et al. Solutions for a cultivated planet. *Nature*, 478(7369):337–342, 2011.

[16] Anthony Fuller, Koreen Millard, and James Green. Croma: Remote sensing representations with contrastive radar-optical masked autoencoders. *Advances in Neural Information Processing Systems*, 36:5506–5538, 2023.

[17] Bo-Cai Gao. Ndwi—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment*, 58(3):257–266, 1996.

[18] Anatol Garioud, Nicolas Gonthier, Loic Landrieu, Apolline De Wit, Marion Valette, Marc Poupée, Sébastien Giordano, et al. Flair: a country-scale land cover semantic segmentation dataset from multi-source optical imagery. *Advances in Neural Information Processing Systems*, 36:16456–16482, 2023.

[19] Sebastian Gerard, Yu Zhao, and Josephine Sullivan. Wildfirespreadts: A dataset of multi-modal time series for wildfire spread prediction. *Advances in Neural Information Processing Systems*, 36:74515–74529, 2023.

[20] Anatoly A Gitelson. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of plant physiology*, 161(2):165–173, 2004.

[21] Abdurrahman Gonenc, Mehmet Sirac Ozerdem, and ACAR Emrullah. Comparison of ndvi and rvi vegetation indices using satellite images. In *2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, pages 1–4. IEEE, 2019.

[22] Xin Guo, Jiangwei Lao, Bo Dang, Yingying Zhang, Lei Yu, Lixiang Ru, Liheng Zhong, Ziyuan Huang, Kang Wu, Dingxiang Hu, et al. Skysense: A multi-modal remote sensing foundation model towards universal interpretation for earth observation imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27672–27683, 2024.

[23] Cassandra Handan-Nader and Daniel E Ho. Deep learning to map concentrated animal feeding operations. *Nature Sustainability*, 2(4):298–306, 2019.

[24] Zhenchun Hao, Sichun Chen, Zehua Li, Zhongbo Yu, Quanxi Shao, Fei Yuan, and Fangxin Shi. Quantitative assessment of the impacts of irrigation on surface water fluxes in the tarim river, china. *Hydrology Research*, 46(6):996–1007, 2015.

[25] Oishee Bintey Hoque, Samarth Swarup, Abhijin Adiga, Sayjro Kossi Nouwakpo, and Madhav Marathe. Irrnet: Advancing irrigation mapping with incremental patch size training on remote sensing imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5460–5469, 2024.

[26] Oishee Bintey Hoque, Abhijin Adiga, Aniruddha Adiga, Siddharth Chaudhary, Madhav Marathe, SS Ravi, Kirti Rajagopalan, Mandy Wilson, and Samarth Swarup. IGraSS: Learning to Identify Infrastructure Networks from Satellite Imagery by Iterative Graph-constrained Semantic Segmentation. In *International Joint Conference on Artificial Intelligence (IJCAI-25)*, 2025.

[27] Oishee Bintey Hoque, Nibir Chandra Mandal, Abhijin Adiga, Samarth Swarup, Sayjro Kossi Nouwakpo, Mandy Wilson, SS Ravi, and Madhav Marathe. Knowledge-Informed Deep Learning for Irrigation Type Mapping from Remote Sensing. In *International Joint Conference on Artificial Intelligence (IJCAI-25)*, 2025.

[28] Alfredo Huete, Kamel Didan, Tomoaki Miura, E Patricia Rodriguez, Xiang Gao, and Laerte G Ferreira. Overview of the radiometric and biophysical performance of the modis vegetation indices. *Remote sensing of environment*, 83(1-2):195–213, 2002.

[29] Calvin Hung, Zhe Xu, and Salah Sukkarieh. Feature learning based approach for weed classification using high resolution aerial images from a digital camera mounted on a uav. *Remote Sensing*, 6(12):12037–12054, 2014.

[30] Jeremy Andrew Irvin, Emily Ruoyu Liu, Joyce Chuyi Chen, Ines Dormoy, Jinyoung Kim, Samar Khanna, Zhuo Zheng, and Stefano Ermon. Teochat: A large vision-language assistant for temporal earth observation data. *arXiv preprint arXiv:2410.06234*, 2024.

[31] Xiaowei Jia, Ankush Khandelwal, David J Mulla, Philip G Pardey, and Vipin Kumar. Bringing automated, remote-sensed, machine learning methods to monitoring crop landscapes at scale. *Agricultural Economics*, 50:41–50, 2019.

[32] Xiaowei Jia, Mengdie Wang, Ankush Khandelwal, Anuj Karpatne, and Vipin Kumar. Recurrent Generative Networks for Multi-Resolution Satellite Data: An Application in Cropland Monitoring. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*, pages 2628–2634, 2019. doi: 10.24963/ijcai.2019/365.

[33] David Ketchum, Kelsey Jencso, Marco P Maneta, Forrest Melton, Matthew O Jones, and Justin Huntington. Irrmapper: A machine learning approach for high resolution mapping of irrigated agriculture across the western us. *Remote Sensing*, 12(14):2328, 2020.

[34] Kartik Kuckreja, Muhammad Sohail Danish, Muzammal Naseer, Abhijit Das, Salman Khan, and Fahad Shahbaz Khan. Geochat: Grounded large vision-language model for remote sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27831–27840, 2024.

[35] Nataliia Kussul, Mykola Lavreniuk, Sergii Skakun, and Andrii Shelestov. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5):778–782, 2017.

[36] Les Levidow, Daniele Zaccaria, Rodrigo Maia, Eduardo Vivas, Mladen Todorovic, and Alessandra Scardigno. Improving water-efficient irrigation: Prospects and difficulties of innovative practices. *Agricultural Water Management*, 146:84–94, 2014. ISSN 0378-3774. doi: https://doi.org/10.1016/j.agwat.2014.07.012. URL `https://www.sciencedirect.com/science/article/pii/S037837741400211X`.

[37] Joel Lisonbee, Molly Woloszyn, and Marina Skumanich. Making sense of flash drought: definitions, indicators, and where we go from here. *Journal of Applied and Service Climatology*, 2021(1):1–19, February 2021. ISSN 2643-0223. doi: 10.46275/joasc.2021.02.001.

[38] Fan Liu, Delong Chen, Zhangqingyun Guan, Xiaocong Zhou, Jiale Zhu, Qiaolin Ye, Liyong Fu, and Jun Zhou. Remoteclip: A vision language foundation model for remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[39] Shengzhong Liu, Tomoyoshi Kimura, Dongxin Liu, Ruijie Wang, Jinyang Li, Suhas Diggavi, Mani Srivastava, and Tarek Abdelzaher. Focal: Contrastive learning for multimodal time-series sensing signals in factorized orthogonal latent space. *Advances in Neural Information Processing Systems*, 36:47309–47338, 2023.

[40] Thomas R Loveland, Bradley C Reed, Jesslyn F Brown, Donald O Ohlen, Zhiliang Zhu, LWMJ Yang, and James W Merchant. Development of a global land cover characteristics database and igbp discover from 1 km avhrr data. *International journal of remote sensing*, 21(6-7): 1303–1330, 2000.

[41] Jialin Luo, Yuanzhi Wang, Ziqi Gu, Yide Qiu, Shuaizhen Yao, Fuyun Wang, Chunyan Xu, Wenhua Zhang, Dan Wang, and Zhen Cui. Mmm-rs: A multi-modal, multi-gsd, multi-scene remote sensing dataset and benchmark for text-to-image generation. *arXiv preprint arXiv:2410.22362*, 2024.

[42] Keyvan Malek, Jennifer Adam, Claudio Stockle, Michael Brady, and Kirti Rajagopalan. When should irrigators invest in more water-efficient technologies as an adaptation to climate change? *Water Resources Research*, 54(11):8999–9032, 2018.

 https://doi.org/10.52202/085713-0186

[43] Nibir Chandra Mandal, Oishee Bintey Hoque, Abhijin Adiga, Samarth Swarup, Mandy Wilson, Lu Feng, Yangfeng Ji, Miaomiao Zhang, Geoffrey Fox, and Madhav Marathe. IrrMap: A Large-Scale Comprehensive Dataset for Irrigation Method Mapping. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD-25)*, 2025.

[44] S Meivel and SJMS Maheswari. Monitoring of potato crops based on multispectral image feature extraction with vegetation indices. *Multidimensional Systems and Signal Processing*, 33 (2):683–709, 2022.

[45] Andres Milioto, Philipp Lottes, and Cyrill Stachniss. Real-time blob-wise sugar beets vs weeds classification for monitoring fields using convolutional neural networks. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4:41–48, 2017.

[46] Anders Krogh Mortensen, Mads Dyrmann, Henrik Karstoft, Rasmus Nyholm Jørgensen, and René Gislum. Semantic segmentation of mixed crops using deep convolutional neural network. 2016.

[47] Kudzai S Mpakairi, Timothy Dube, Mbulisi Sibanda, and Onisimo Mutanga. Fine-scale characterization of irrigated and rainfed croplands at national scale using multi-source data, random forest, and deep learning algorithms. *ISPRS Journal of Photogrammetry and Remote Sensing*, 204:117–130, 2023.

[48] Multi-Resolution Land Characteristics Consortium (MRLC). National land cover database (nlcd), 2021. URL `https://www.mrlc.gov/`. Accessed: 2024-03-01.

[49] Robert Mueller and Jeffrey Harris. Reported uses of cropscape and the national cropland data layer program. In *Proceedings of the Sixth International Conference on Agricultural Statistics (ICAS VI)*, Rio de Janeiro, Brazil, October 23–25 2013. Posted January 31, 2014.

[50] Vishal Nedungadi, Ankit Kariryaa, Stefan Oehmcke, Serge Belongie, Christian Igel, and Nico Lang. Mmearth: Exploring multi-modal pretext tasks for geospatial representation learning. In *European Conference on Computer Vision*, pages 164–182. Springer, 2024.

[51] Mubashir Noman, Noor Ahsan, Muzammal Naseer, Hisham Cholakkal, Rao Muhammad Anwer, Salman Khan, and Fahad Shahbaz Khan. Cdchat: A large multimodal model for remote sensing change description. *arXiv preprint arXiv:2409.16261*, 2024.

[52] USDA Natural Resources Conservation Service (NRCS). Ssurgo — soil survey geographic database, 2023. URL `https://nrcs.app.box.com/v/soils/folder/233398887779`. Accessed: 2024-03-01.

[53] Mahmoud Osman, Benjamin Zaitchik, Jason Otkin, and Martha Anderson. A global flash drought inventory based on soil moisture volatility. *Scientific Data*, 11(1), September 2024. ISSN 2052-4463. doi: 10.1038/s41597-024-03809-9.

[54] Val Osowski. Long-term mapping key to effective management of irrigated areas. `https://msutoday.msu.edu/news/2020/long-term-mapping-key-to-effective-management-of-irrigated-areas`, 2020. Accessed: 2024-03-24.

[55] C Dionisio Pérez-Blanco, Arthur Hrast-Essenfelder, and Chris Perry. Irrigation technology and water conservation: A review of the theory and evidence. *Review of Environmental Economics and Policy*, 2020.

[56] Md Shahriar Pervez and Jesslyn F Brown. Mapping irrigated lands at 250-m scale by merging modis data and national agricultural statistics. *Remote Sensing*, 2(10):2388–2412, 2010.

[57] Sophie Plassin, Jennifer Koch, Madison Wilson, Kevin Neal, Jack R. Friedman, Stephanie Paladino, and James Worden. Multi-scale fallow land dynamics in a water-scarce basin of the u.s. southwest. *Journal of Land Use Science*, 16(3):291–312, May 2021. ISSN 1747-4248. doi: 10.1080/1747423x.2021.1928310.

[58] Yongquan Qu, Juan Nathaniel, Shuolin Li, and Pierre Gentine. Deep generative data assimilation in multimodal setting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 449–459, 2024.

[59] Brian D. Richter, Dominique Bartak, Peter Caldwell, Kyle Frankel Davis, Peter Debaere, Arjen Y. Hoekstra, Tianshu Li, Landon Marston, Ryan McManamay, Mesfin M. Mekonnen, Benjamin L. Ruddell, Richard R. Rushforth, and Tara J. Troy. Water scarcity and fish imperilment driven by beef production. *Nature Sustainability*, 3(4):319–328, March 2020. ISSN 2398-9629. doi: 10.1038/s41893-020-0483-z.

[60] Caleb Robinson, Ben Chugg, Brandon Anderson, Juan M. Lavista Ferres, and Daniel E. Ho. Mapping industrial poultry operations at scale with deep learning and aerial imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:7458–7471, 2022. doi: 10.1109/JSTARS.2022.3191544.

[61] Philip Sedgwick. Cohen's coefficient κ. *BMJ*, 344, 2012.

[62] Stefan Siebert, Jacob Burke, Jean-Marc Faures, Karen Frenken, Jippe Hoogeveen, Petra Döll, and Felix Theodor Portmann. Groundwater use for irrigation–a global inventory. *Hydrology and earth system sciences*, 14(10):1863–1880, 2010.

[63] Stefan Siebert, Verena Henrich, Karen Frenken, and Jacob Burke. Update of the digital global map of irrigation areas to version 5. *Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany and Food and Agriculture Organization of the United Nations, Rome, Italy*, 10(2.1):2660–6728, 2013.

[64] Shivangi Srivastava, John E Vargas-Munoz, and Devis Tuia. Understanding urban landuse from the above and ground perspectives: A deep learning, multimodal solution. *Remote sensing of environment*, 228:129–143, 2019.

[65] Pardhasaradhi Teluguntla, Prasad S Thenkabail, Jun Xiong, Murali Krishna Gumma, Chandra Giri, Cristina Milesi, Mutlu Ozdogan, Russ Congalton, James Tilton, Temuulen Tsagaan Sankey, et al. Global cropland area database (gcad) derived from remote sensing in support of food security in the twenty-first century: current achievements and future possibilities. 2015.

[66] Prasad S Thenkabail, Chandrashekhar M Biradar, Praveen Noojipady, Venkateswarlu Dheeravath, Yuanjie Li, Manohar Velpuri, Muralikrishna Gumma, Obi Reddy P Gangalakunta, Hugh Turral, Xueliang Cai, et al. Global irrigated area map (giam), derived from remote sensing, for the end of the last millennium. *International journal of remote sensing*, 30(14):3679–3733, 2009.

[67] Franck Thénot, Maurice Méthy, and Thierry Winkel. The photochemical reflectance index (pri) as a water-stress index. *International Journal of Remote Sensing*, 23(23):5135–5139, 2002.

[68] Compton J Tucker. Satellite remote sensing of total herbaceous biomass production in the senegalese sahel: 1980–1984. *Remote sensing of environment*, 17(3):233–249, 1985.

[69] United States Geological Survey. Verified-irrigated-agricultural-lands-for-the-United-States. `https://catalog.data.gov/dataset/verified-irrigated-agricultural-lands-for-the-united-states-200217`, year = 2025, note = [Accessed 03-2025],.

[70] USCB. Tiger/line shapefiles. `https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html`, 2025. [Accessed 03-2025].

[71] USDA. Usda nass cropland data layers. `https://developers.google.com/earth-engine/datasets/catalog/USDA_NASS_CDL`, 2025. [Accessed 03-2025].

[72] USDA-NWIS. Water data for the nation. `https://waterdata.usgs.gov/nwis?`, 2025. [Accessed 03-2025].

[73] U.S. Geological Survey (USGS). Fews net monthly evapotranspiration, 2023. URL `https://earlywarning.usgs.gov/fews/product/460/`. Accessed: 2024-03-01.

 https://doi.org/10.52202/085713-0186

[74] UTAH-DNR. Water-related land use (wrlu). https://dwre-utahdnr.opendata.arcgis.com/pages/wrlu-data, 2025. [Accessed 03-2025].

[75] Shyamal S Virnodkar, Vinod K Pachghare, Virupakshagouda C Patil, and Sunil Kumar Jha. Denseresunet: An architecture to assess water-stressed sugarcane crops from sentinel-2 satellite imagery. *Traitement du Signal*, 38(4), 2021.

[76] Marie Weiss, Frédéric Jacob, and Grgory Duveiller. Remote sensing for agricultural applications: A meta-review. *Remote sensing of environment*, 236:111402, 2020.

[77] WSDA. Washington state department of agriculture agricultural land use dataset. https://agr.wa.gov/departments/land-and-water/natural-resources/agricultural-land-use, 2025. [Accessed 03-2025].

[78] Yanhua Xie, Holly K Gibbs, and Tyler J Lark. Landsat-based irrigation dataset (lanid): 30 m resolution maps of irrigation distribution, frequency, and change for the us, 1997–2017. *Earth System Science Data*, 13(12):5689–5710, 2021.

[79] Kun Zhang, Xin Li, Donghai Zheng, Ling Zhang, and Gaofeng Zhu. Estimation of global irrigation water use by the integration of multiple satellite observations. *Water Resources Research*, 58(3):e2021WR030031, 2022.

[80] Baojuan Zheng, James B Campbell, and Kirsten M de Beurs. Remote sensing of crop residue cover using multi-temporal landsat imagery. *Remote Sensing of Environment*, 156:312–321, 2014.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: This paper contributes a novel spatial dataset and an automated pipeline to generate the same for future extensions to the data. These are summarized in the abstract and explained in the introduction with background and motivation.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: A limitations section has been added.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

https://doi.org/10.52202/085713-0186

Answer: [NA]

Justification: [NA]

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper provides full description and details in the main part and the supplement. Code and data are made open.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Code and data have been released.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: All the details of training and model tuning are provided in the experiment setup and the supplement.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [NA]

Justification: Standard accuracy, F1-score metrics provided.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.

https://doi.org/10.52202/085713-0186

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Compute environment is described.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The paper uses open datasets. No research involving human subjects or participants conducted.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The paper describes the various potential beneficial applications of the dataset for the domain as well as development of novel methods.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: A static dataset is contributed.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The papers and websites have been cited.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

https://doi.org/10.52202/085713-0186

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: Documentation has been provided.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: [NA]

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: Publicly available data sources used.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [Yes]

Justification: Visual-LLM methods are applied in the pipeline for semi-automated data processing. The exact methodology, inlcuding applied textual prompts, has been documented. Some portions of the implementation (e.g., data preprocessing, visualization scripts) were assisted using large language models (LLMs) such as ChatGPT for coding support. All code was reviewed and validated by the authors.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.

https://doi.org/10.52202/085713-0186