TIRAuxCloud: A Thermal Infrared Dataset for Day and Night Cloud Detection

Alexis Apostolakis[1,3], Vasileios Botsos[1,2], Niklas Wölki[4], Andrea Spichtinger[4], Nikolaos Ioannis Bountos[2], Ioannis Papoutsis[2,3], and Panayiotis Tsanakas[1]

[1]National Technical University of Athens, School of Electrical and Computer Engineering
[2]School of Rural, Surveying and Geoinformatics Engineering, National Technical University of Athens
[3]National Observatory of Athens
[4]OroraTech GmbH

***Abstract*—Clouds are a major obstacle in Earth observation, limiting the usability and reliability of critical remote sensing applications such as fire disaster response, urban heat island monitoring, and snow and ice cover mapping. Therefore, the ability to detect clouds 24/7 is of paramount importance. While visible and near-infrared bands are effective for daytime cloud detection, their dependence on solar illumination makes them unsuitable for nighttime monitoring. In contrast, thermal infrared (TIR) imagery plays a crucial role in detecting clouds at night, when sunlight is absent. Due to their generally lower temperatures, clouds emit distinct thermal signatures that are detectable in TIR bands. Despite this, accurate nighttime cloud detection remains challenging due to limited spectral information and the typically lower spatial resolution of TIR imagery. To address these challenges, we present *TIRAuxCloud*, a multimodal dataset centered around thermal spectral data to facilitate cloud segmentation under both daytime and nighttime conditions. The dataset comprises a unique combination of multispectral data (TIR, optical, and near-infrared bands) from Landsat and VIIRS, aligned with auxiliary information layers. Elevation, land cover, meteorological variables, and cloud-free reference images are included to help reduce surface-cloud ambiguity and cloud formation uncertainty. To overcome the scarcity of manual cloud labels, we include a large set of samples with automated cloud masks and a smaller manually annotated subset to further evaluate and improve models. Comprehensive benchmarks are presented to establish performance baselines through supervised and transfer learning, demonstrating the dataset's value in advancing the development of innovative methods for day and night time cloud detection.**

***Index Terms*—Thermal Infrared, Cloud Detection, Nighttime Remote Sensing, Multimodal Fusion, Semantic Segmentation, Transfer Learning.**

I. Introduction

Accurate and consistent cloud detection remains a fundamental requirement for a wide range of remote sensing applications not only during the day but during the night as well. The presence of clouds acts as a major obstacle to satellite-based observation, as they obscure the retrieval of surface and atmospheric information [1]–[5]. As a result, several critical remote sensing services could underperform or send misleading signals to end users, causing them to act (or not act) on faulty data without being aware of it. Notable examples of remote sensing services that are needed to operate day and night are *fire detection, monitoring and assessment*, *precipitation forecasting*, *urban heat island*, *surface temperature*, *snow and ice cover mapping*, *aerosol optical depth* [1], [6]–[8]. While cloud detection methods have seen significant advancements, particularly for daytime remote sensing images, nighttime cloud detection presents substantial challenges [2], [4], [9]–[11]. During the day, visible and near infrared (VNIR) bands are effective due to the high solar reflectivity of clouds. However, the absence of direct solar illumination at night makes these bands unsuitable [9], [10]. On the other hand, thermal infrared (TIR) radiation is continuously emitted and can be observed regardless of solar illumination using satellite TIR sensors from satellite missions such as Landsat, VIIRS, and MODIS [12], [13], or even with ground-based TIR sensors [3]. Clouds emit thermal energy proportional to their temperature, and they are generally colder than the Earth's surface or ocean temperatures. This temperature difference allows for their detection using thermal imagery from satellites in wavelengths such as the 8.0-12.0 μm window region (LWIR) [4]. A significant challenge that arises when relying on thermal infrared data is distinguishing clouds from snow and ice, because both are cold and have similar temperatures [1]. Existing methods (either based on traditional thresholding [14]–[16] or more advanced deep learning approaches [3], [10], [17]–[19]) struggle in night-time scenarios.

A significant obstacle hindering accurate all-day cloud detection has been the absence of large, publicly available datasets specifically designed for cloud detection, using data available during day and night such as thermal emission observations [8], [10], [11]. However, cloud properties are controlled by a range of environmental factors that are not fully captured in the thermal signal alone. Some studies have shown that incorporation of auxiliary information, such as Digital Elevation Model (DEM), Land Use Land Cover (LULC) or meteorological values, can provide valuable contextual cues for distinguishing clouds from cold surfaces such as snow, ice or high elevation terrain. In particular, DEM has shown to improve cloud and snow separation in mountainous regions [1], [20] and atmospheric state variables affect cloud formation and thickness [21]. Despite these indications, current public datasets rarely include such auxiliary layers and are typically limited to spectral imagery alone, as seen in (Table I).

To address these gaps, we introduce *TIRAuxCloud*, a novel dataset for cloud detection covering day and night conditions, primarily based on thermal infrared data. TIRAuxCloud com-

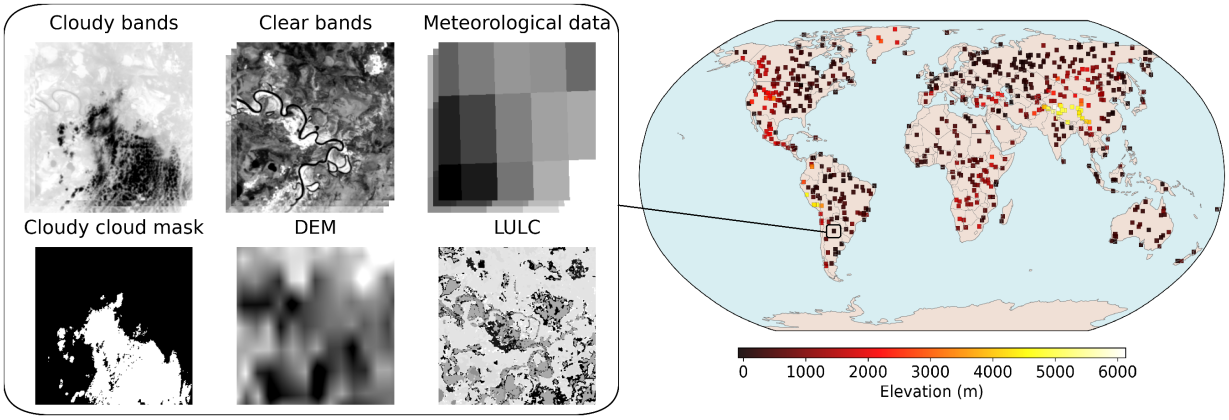

Fig. 1. Global distribution of sampled Landsat scenes used for training, colored by elevation. Each one contains all the cloudy bands mentioned in Section II alongside their clear counterparts and the auxiliary bands. Cloudy bands panel displays Cloudy Radiance B10/B11 mean, while the Clear bands panel shows Clear Radiance B10/B11 mean.

bines imagery from established satellite missions, specifically Landsat 8/9 and VIIRS, known for their thermal infrared capabilities. We provide all spectral bands from both satellites enabling studies for day and night conditions. The dataset includes 21,058 Landsat image samples labeled by the automated Landsat algorithm [16], 2,563 images extracted from manually annotated scenes from a USGS dataset [22], [23] and 5,004 VIIRS images created from scenes downloaded from World Bank [24] collections. Our approach extends beyond raw satellite imagery by integrating auxiliary data to investigate the potential to enhance cloud detection accuracy, under conditions where reflectance spectral bands are unavailable. These auxiliary data, chosen for their availability during both day and night, include i) *Cloud-free counterparts* precisely co-registered with cloudy scenes, serving as clear-sky references; ii) *DEM* useful for improving thin or broken cloud detection, especially in complex terrain [11]; iii) *Weather* parameters from Copernicus ERA5 [25] such as temperature, dew point temperature, surface pressure and precipitation that significantly influence cloud formation, cloud types and states [3], [26]; iv) *LULC maps* which help capture how different surfaces emit thermal radiation, aiding cloud detection over challenging backgrounds like snow, bare land, and artificial lights [10], [11]; and finally v) *Metadata* from Landsat and VIIRS such as Geolocation, Acquisition Date and Time, Sun Elevation and Azimuth Angles, which can serve as additional model inputs or support statistical analyses [4], [9], [27], [28]

By providing this rich, multi-modal dataset, we aim to overcome the limitations of current training data for all-day cloud segmentation, paving the way for more robust and accurate deep learning or foundation models capable of performing the task. Our contributions are i) the integration of, to the best of our knowledge, the most comprehensive set of auxiliary data available for day and night cloud detection offering rich contextual information to support learning; ii) the benchmarking with widely adopted architectures and models tailored for cloud detection, establishing strong baselines for future research; iii) the provision of strong indications that auxiliary data can enhance model performance; and iv) evidence that the main automatically annotated Landsat dataset offers sufficiently rich information to support robust model training, even more than the smaller manually annotated Landsat dataset.

A. Related Work

Deep learning has enabled effective cloud segmentation, with many models using encoder-decoder architectures, some tailored to the specific challenges of the task [1], [3], [5], [10], [18], [29]. CDnetV2 [17] and Cloud-Adapter [30] are two notable examples with open-source PyTorch implementations. CDnetV2 focuses on cloud-snow coexistence using a CNN encoder-decoder with adaptive feature fusion and semantic guidance. Cloud-Adapter leverages pre-trained Vision Foundation Models, enhancing segmentation with minimal extra parameters.

Recent work has also focused on the use of thermal infrared data for all-day cloud segmentation. [11] introduces a unique approach to nighttime satellite cloud detection by generating machine learning training samples from daytime thermal infrared data, exploiting the consistent thermal characteristics of clouds across day and night. SwinCloud [10] introduces, for the first time, a Swin-UNet for cloud detection, integrating a Global–Local Swin Transformer block (GLST) and a Feature Fusion Module (FFM) for robust, multi-scale cloud detection in thermal infrared imagery. Other models, such as MFFCD-Net and CloudU-Net variants, address day/night detection challenges using VIIRS Day/Night Band and ground-based TIR camera images as input, respectively [2], [3].

Several publicly available datasets (Table I) support cloud detection and related tasks, mainly relying on reflectance bands (e.g., visible and near-infrared) from satellite imagery. The *38-Cloud* [31] and *95-Cloud* [32] datasets consist of 38 and 95 manually annotated Landsat 8/9 scenes, respectively, providing a total of 70,404 image patches. However, these datasets do not include processed thermal bands from Landsat imagery. The *CloudSEN12* [33] dataset and its extension, *CloudSEN12+* [34], are large, globally distributed collections based primarily on Sentinel-2 imagery, offering 49,400 image patches for the

TABLE I
OVERVIEW OF PUBLICLY AVAILABLE REMOTE SENSING DATASETS FOR CLOUD DETECTION

Dataset	**Source**	**Bands Auxiliary**	**Samples Number**	**Size**	**Resolution** (Resampled)	**Mask**	**Notes**
38/95-Cloud	Landsat 8/9	VNIR	34,701	384px	30 m	Manual	No TIR
CloudSEN12+	Sentinel 1/2	VNIR, SAR, DEM, LULC	50,249	509px, 2000px	10 m	Manual	No TIR
AllClear	Sentinel 1/2, Landsat 8/9	VNIR, SAR, DEM, LULC, Metadata	4,354,652	256px	10 m	Auto	Cloud removal focus; small TIR patches
SPARCS	Landsat 8	TIR, VNIR, Metadata	80	1000px	30 m	Manual	TIR capable; no auxiliary; limited samples
L8 Biome	Landsat 8	VNIR, TIR	192	100,000 m (scenes)	30 m	Manual	TIR capable; no auxiliary; moderate samples
TIRAuxCloud	Landsat 8/9, VIIRS	TIR, VNIR, DEM, Met., LULC, Metadata	28,625	256px	Landsat 100m VIIRS 200m	Auto & Manual	TIR with auxiliary data; sufficient samples

semantic segmentation of clouds and cloud shadows. The dataset includes high-quality manual annotations for thick clouds, thin clouds, and cloud shadows, along with Sentinel-2 spectral bands, co-registered Sentinel-1 SAR imagery, and auxiliary data such as DEMs and LULC maps. However, it lacks thermal infrared imagery, limiting its applicability to daytime conditions only. *AllClear* [28] is the largest public dataset for cloud removal, comprising 23,742 globally distributed regions and over 4 million images. It includes multispectral imagery from Sentinel-1/2 and Landsat-8/9, with thermal bands from Landsat. However, the resampling of all bands to 10m to match Sentinel-2 optical bands results in small patches (2.5km), limiting suitability for cloud segmentation using Landsat thermal data, which has 100m native resolution. *SPARCS* is a widely used public dataset for identifying clouds and cloud shadows, consisting of 80 Landsat-8 scenes, however it is relatively small in size, which limits its suitability for training robust deep learning cloud detection models that rely on thermal information. The *L8 Biome* dataset [14], [22], [23] is one of the few manually annotated resources that include thermal data (Landsat bands 10 and 11), though it is provided as full scenes requiring preprocessing for model training. It is also limited in size, probably reducing the model's ability to generalize.

Landsat 8/9 and several other satellite missions contribute thermal infrared (TIR) data for cloud detection under both day and night conditions. Landsat's automated cloud masks are generated using the pixel-based Fmask algorithm, for which up to 96% accuracy has been reported [16]. VIIRS (Visible Infrared Imaging Radiometer Suite) offers 16 moderate-resolution channels used in neural network-based cloud masks (e.g., NNCM), along with VIIRS/CrIS fusion-derived features approximating MODIS-like channels. SDGSAT-1, via its Thermal Infrared Imager (TIS), provides three TIR bands (B1–B3) and is commonly used to evaluate model transferability across sensors. MODIS and Himawari-8 supply multispectral coverage, including key TIR wavelengths (8–14μm), enabling reliable cloud detection in nighttime imagery. Finally, AVHRR/HRPT supports day-night cloud detection using infrared channels 3–5, without reliance on visible light, making it effective under low-illumination conditions.

II. DATASET DESCRIPTION

To support the development of accurate cloud detection models capable of operating under varying illumination and surface conditions, we construct TIRAuxCloud[1], a dataset compiled from multiple satellite sources, centered on thermal infrared (TIR) observations and enriched with auxiliary geospatial layers. The dataset is designed to address limitations in existing resources by providing structured, spatially aligned data suitable for both daytime and nighttime segmentation tasks. The dataset is organized into three distinct subsets: the *Landsat Main* Dataset, the Manually Annotated (MA) Landsat Dataset, and the VIIRS Dataset. By integrating thermal observations with auxiliary surface and atmospheric information, TIRAuxCloud supports robust cloud segmentation across a wide range of geographic regions, temporal variations, and illumination conditions.

A. Landsat Main Dataset

The *Landsat Main* subset is the primary component of TIRAuxCloud. It is designed to provide a reliable and representative foundation for supervised cloud segmentation. Leveraging thermal infrared channels in conjunction with reflective bands and geospatial context, this subset supports learning under diverse surface, atmospheric, and illumination conditions. A key objective in constructing this set was to provide Landsat sensor data, alongside high quality auxiliary layers. This enables TIR-based models to better distinguish clouds from cold surfaces such as snow/ice or cold mountain tops, particularly during nighttime. Similarly, it can help VNIR-based models distinguish bright surfaces (e.g. snow, urban areas) from clouds during the day. Each cloudy scene is paired with a clear counterpart, providing a cloud-free surface reference that helps distinguish ambiguous or semi-transparent clouds.

All *Landsat Main* scenes are sourced from Landsat 8 and Landsat 9 Collection 2 Level-1 products, accessed through the USGS/EROS M2M API [35]. These products provide radiometrically calibrated surface observations, along with

[1]https://github.com/Orion-AI-Lab/TIRAuxCloud, https://huggingface.co/datasets/tirauxcloud/TIRAuxCloud

pixel level quality information (QA PIXEL band). To construct a representative and balanced training set, we employed a stratified sampling strategy, the results of which are illustrated in Figures 1, 3 and 4. Scenes were selected according to three main criteria: spatial diversity, temporal diversity, and cloud coverage. To ensure spatial diversity, land locations were randomly selected from a global coverage grid, ensuring balanced distribution across continents, climate zones, and elevation ranges. Temporal diversity was achieved by selecting scenes distributed approximately uniformly across all twelve months, capturing seasonal variations in vegetation, snow presence, solar angle, and atmospheric conditions. We focus on scenes with total reported cloud cover between 30 and 70%, prioritizing partially cloudy conditions. For each selected cloudy scene, we retrieved a corresponding "clear" scene from the same location, with reported cloud cover below 5%, captured as close in time as possible, typically within 30 days, to serve as a cloud-free reference. The final sampling resulted in 617 cloudy and clear scene pairs, from which a total of 21,058 patches were extracted. In addition to the satellite observations, we augment each cloudy scene with a set of auxiliary geospatial layers that capture terrain, surface type, and atmospheric context. Specifically, we extract a DEM derived from the NOAA Global Relief Model [36] and a LULC layer informed by the Copernicus Global Land Cover 100m dataset [37]. Finally, we include five cloud related meteorological variables (surface pressure, dew point temperature, skin temperature, total precipitation and snow cover), obtained by ERA5 Land data [38]. Precipitation and snow cover values are log-transformed to reduce distribution skewness, facilitating normalization and stabilizing model training. All auxiliary bands are spatially aligned with the cloudy scene observation to ensure consistent pixel-level correspondence. These layers are included to provide additional spatial and environmental context for cloud interpretation.

All patches are resampled to a fixed resolution of 100 metres and cropped to 256×256 pixels, using average resampling, except for the weather, elevation and LULC layers which were resampled using nearest neighbor. Cloud masks are generated using the information in the QA PIXEL band. Pixels are classified as clear, thin cloud, or cloud based on the high confidence flags provided in QA PIXEL. To ensure a clean and processing-ready dataset, patches containing missing or invalid values in any spectral band or auxiliary layer were excluded during patch extraction. Each one of the patches contains 39 bands. The 16 bands from the cloudy scene include Landsat bands B1–B11, spanning reflective and thermal infrared wavelengths, the QA PIXEL band, a cloud mask derived from QA PIXEL, the radiance values of Bands 10 and 11, and the mean radiance across Bands 10 and 11. The clear counterpart includes an identical set of 16 bands, allowing direct comparison with the corresponding cloudy scene. In addition, each patch includes DEM, LULC and the five meteorological bands described previously.

Figure 2 shows example bands from a 256×256 patch, including cloudy and clear radiance (B10/B11), RGB composites, the corresponding cloud mask, and auxiliary layers such as DEM, LULC, surface pressure, and skin temperature.

B. MA Landsat Dataset

The *MA Landsat* dataset is a complementary subset derived through patch extraction from the manually annotated cloud mask dataset described in [22], [23]. It primarily serves as an evaluation set, providing high-quality manual annotations for assessing model performance, and it can also be used for enhancing model performance. Compared to the *Landsat Main* dataset, it is approximately an order of magnitude smaller.

The cloud annotations are sourced from the Landsat 8 Collection 1 and Collection 2 Cloud Truth Mask Validation Set, a human labeled dataset developed by USGS for the purpose of validating cloud detection algorithms [22], [23]. These annotations provide high precision cloud masks with pixel level labels for clear, thin cloud, and cloud classes. This subset is composed of 68 Landsat 8 scenes, from which we extract a total of 2,563 patches. In addition to the satellite observations, each patch is supplemented with auxiliary layers that mirror those used in the *Landsat Main* set, including a DEM, a LULC map, and five meteorological variables. All auxiliary layers are spatially aligned with the annotated cloudy observation to ensure consistent and reliable context for model evaluation. Each patch in the *MA Landsat* subset has the same size (256x256) resolution (100m) and includes the same 39 bands as those used in the Main Landsat Dataset, ensuring consistency in the input format for model training.

C. VIIRS Dataset

The VIIRS Dataset is constructed to assess cloud detection model performance under nighttime imaging conditions, where visible reflectance is largely absent and thermal infrared observations become the primary source of information. This subset complements the daytime Landsat-based evaluation by specifically targeting challenging low-illumination scenarios. By doing so, it enables the evaluation of cloud segmentation models in environments where traditional optical signals are unavailable or unreliable, emphasizing robustness to lighting conditions.

To ensure broad temporal representation and geographic diversity, we select twelve orbital scenes from the Visible Infrared Imaging Radiometer Suite (VIIRS), one for each month of the year (Figure 4). These scenes are obtained from the Light Every Night (LEN) project, a public dataset compiled by the World Bank that provides daily VIIRS nighttime observations [24]. Scenes are chosen to reflect a range of global environments, from urban to rural, tropical to polar, and across various atmospheric states. From these twelve segments, we extract a total of 5,004 image patches, providing a substantial and seasonally distributed test sample for model benchmarking. Each patch contains five observational bands obtained directly from the LEN dataset. These include lunar illuminance (li), visible imagery (vis), thermal infrared brightness (tir), pixel sample position (samples), and a 16-bit bitmask (flag) that encodes quality and cloud information. In addition, each patch includes a derived binary cloud mask band, constructed by interpreting the flag bitmask and labeling pixels as either clear or cloud based on high confidence cloud indicators.

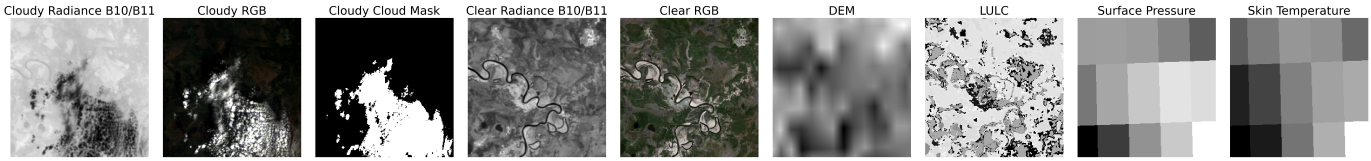

Fig. 2. Example bands from a 256×256 Landsat 8 patch.

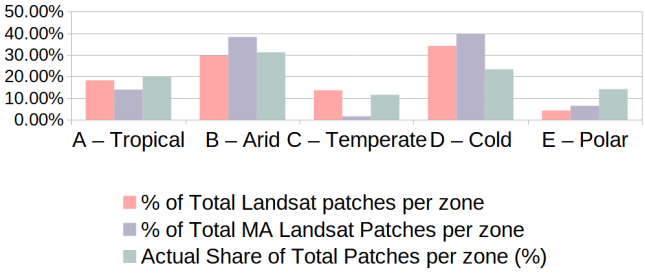

Fig. 3. Distribution of Landsat and Manually Annotated Landsat patches across Köppen–Geiger climate zones [39].

All patches are resampled to a uniform spatial resolution of 200 metres and cropped to a fixed size of 256 × 256 pixels. Following the same quality control procedure applied in the other two subsets, we exclude patches that contain missing values in any of the observational bands or that lack valid cloud mask information. Each patch in this subset is thus a valid nighttime observation, ready for use as input for model training and evaluation.

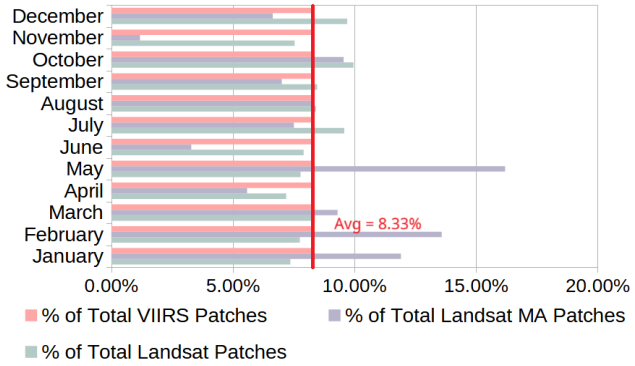

Fig. 4. Temporal distribution of patches across months for VIIRS, Manually Annotated Landsat, and Landsat subsets.

III. Benchmark and Results

To demonstrate the utility of the proposed dataset, we design a comprehensive benchmark, assessing i) the potential of auxiliary features to enhance segmentation capabilities; and (ii) the dataset's capacity to support the extraction of deep, cross-sensor / cross-domain features improving model transferability. Moreover, we offer qualitative results, showcasing how these features can enhance segmentation performance.

A. Feature Combination Benchmarks

In Tables II and III, we present a series of experiments using different feature combinations from the Landsat Main dataset. From the first to the last row, we first show results from models trained with all relevant spectral Landsat bands known to contribute to cloud detection. Subsequently, we evaluate performance using only the thermal radiance band (TIR) and then progressively introduce combinations of the thermal band with various auxiliary inputs: the co-registered clear-sky reference (TIR/Clear), surface elevation (TIR/DEM), meteorological variables (TIR/Meteo), and finally all auxiliary data combined. The Landsat-based training dataset described in Section II is split into 70% for training, 15% for validation, and 15% for testing. A strict group-split policy is applied, where groups correspond to scene IDs, to ensure that no patch from the same scene appears in different splits, thus mitigating spatial correlation. We employ four architectures to represent both generic and cloud-specialized semantic segmentation models. From the convolutional architecture family, we use the widely adopted *U-Net* with a *ResNet34* encoder [40] and the cloud-specific *CDNetV2* [17], a model tailored for cloud segmentation that incorporates channel and spatial attention mechanisms (AFFM) to enhance feature representations (see Section I-A). From the transformers family, we apply *SegFormer* with a *MiT-B2* encoder [41] and *SwinCloud*, a Swin-UNet adaptation for cloud detection that integrates a Global–Local Swin Transformer block (GLST) and a Feature Fusion Module (FFM) for multi-scale cloud detection in thermal infrared imagery (see Section I-A).

We present the Accuracy and mean Intersection over Union (mIoU) in Tables II and III for the two-class (0: clear, 1: cloud) and three-class (0: clear, 1: thin cloud, 2: cloud) targets, respectively. Notably, thin clouds exhibit inherent ambiguity due to their semi-transparent nature, introducing aleatoric label uncertainty. This uncertainty hinders model training by generating inconsistent supervisory signals, thereby degrading performance [42]. As a result, mIoU scores for the three-class problem are consistently lower than those for the two-class setup across all related works. Our experiments indicate that certain feature combinations demonstrate improved performance (bold numbers in tables II, III) over the thermal-only baseline, particularly when using the attention-equipped SegFormer, CDNetV2 and SwinCloud. In contrast, the U-Net model does not appear to benefit from additional inputs. This tendency becomes more pronounced in the 3-class problem and in the combinations where the meteorology variables are included.

TABLE II
MODEL PERFORMANCE (2-CLASS CLOUD SEGMENTATION) ACROSS FEATURE COMBINATIONS

Features	CDnetV2		SegFormer		U-net		SwinCloud	
	Acc	**mIoU**	**Acc**	**mIoU**	**Acc**	**mIoU**	**Acc**	**mIoU**
All Bands	0.93	0.86	0.93	0.86	0.93	0.86	0.94	0.89
TIR	0.87	0.75	0.90	0.81	0.90	0.81	0.89	0.79
TIR/Clear	0.87	**0.76**	0.90	0.81	0.88	0.78	0.86	0.75
TIR/DEM	0.86	0.74	0.90	0.81	0.90	0.80	0.86	0.74
TIR/Meteo	**0.89**	**0.79**	0.90	0.81	0.90	0.80	**0.90**	**0.80**
TIR/Aux.	**0.89**	**0.79**	0.90	0.81	0.89	0.79	**0.90**	**0.80**

TABLE III
MODEL PERFORMANCE (3-CLASS CLOUD SEGMENTATION) ACROSS FEATURE COMBINATIONS

Features	CDnetV2		SegFormer		U-net		SwinCloud	
	Acc	**mIoU**	**Acc**	**mIoU**	**Acc**	**mIoU**	**Acc**	**mIoU**
All Bands	0.91	0.71	0.89	0.66	0.89	0.65	0.90	0.70
TIR	0.83	0.53	0.86	0.56	0.86	0.58	0.80	0.52
TIR/Clear	**0.84**	0.53	0.86	**0.59**	0.84	0.55	0.80	0.50
TIR/DEM	0.83	0.51	0.84	**0.57**	0.85	0.56	0.79	0.49
TIR/Meteo	**0.87**	**0.54**	**0.88**	**0.60**	0.86	0.55	**0.83**	**0.55**
TIR/Aux.	**0.86**	**0.54**	**0.87**	**0.57**	0.86	0.57	**0.86**	**0.55**

B. Visual examples

The following examples illustrate how different auxiliary inputs resolve specific typical errors of TIR-only cloud detection. By contrasting the model's outputs from cloudy TIR alone against those produced when clear TIR, DEM, and meteorological auxiliary data are used, we highlight how each auxiliary source provides complementary information that helps the model avoid common confusions.

In Figure 5a, the clear image reveals a part of a mountainous area where the model trained only on Cloudy TIR incorrectly detects a cloud. The additional input provided to the model trained with both Cloudy and Clear TIR contributes to recognizing the clear ground, whereas the Cloudy TIR-only model fails to distinguish between the lower thermal radiance and the presence of cloud structures.

In Figure 5b, the inclusion of the DEM helped reduce confusion between clouds and elevated terrain. As seen in the related image, the prediction made by using only the TIR band incorrectly labels part of a high mountain region as cloud. This is a common challenge in infrared based cloud detection, where cold surface temperatures at high altitudes can mimic the radiative signature of cloud tops. Once DEM is incorporated, the model correctly identifies this area as terrain rather than cloud, significantly improving prediction accuracy.

The combination of low precipitation and high surface pressure, as seen in Figures 5c and 5d, can play a crucial role in improving the model's ability to differentiate thin clouds from thicker, high confidence cloud cover. High surface pressure is typically associated with stable atmospheric conditions, which inhibits the formation of thick clouds. At the same time, low precipitation levels further support the absence of deep moist convection, favoring the presence of thinner cloud layers.

C. Cross-Domain transferability

To demonstrate the utility of the Landsat Main dataset as a source domain, we perform transfer learning experiments by adapting the learned representations to the MA Landsat and VIIRS datasets. In the case of the *MA Landsat* dataset, the target masks differ slightly because they are manually annotated, resulting in a minor domain shift. In contrast, the *VIIRS* dataset exhibits a much stronger domain shift, as its TIR band has an original spatial resolution of 750 m (compared to 100 m for Landsat) and originates from a different sensor with slightly shifted spectral wavelengths. Table IV showcases the results of the experiments.

Following the order of rows in Table IV, models were first trained using the *U-Net* and *SegFormer* architectures directly on *VIIRS* and *MA Landsat* datasets to establish baselines. Next, we used the models pretrained on *Landsat Main* to perform a series of transfer learning experiments. We ran inference with the pretrained models and then fine-tuned them on *VIIRS* and *MA Landsat* dataset by fully or partially freezing the encoder. In the partial-freeze case, the first three layers of the *U-Net* and the overlapping patch embedding stages of the *SegFormer* were frozen. All models were evaluated on the test splits of the *VIIRS* and *MA Landsat* datasets.

The results in Table IV show that, across all transfer learning experiments, even with no fine-tuning, models pretrained on *Landsat Main* consistently outperform those trained directly on *MA Landsat*. In the case of *VIIRS*, only the SegFormer fine-tuned models with partially frozen encoders pretrained on *Landsat Main* surpass the performance of models trained directly on *VIIRS*.

These findings demonstrate that the *Landsat Main* dataset serves as a strong source domain for learning transferable representations to *MA Landsat*, which has more accurate labels. An important observation is that models pretrained on Landsat Main achieve better performance even without fine-tuning. However, transferring to a different sensor requires substantial fine-tuning, which nevertheless tends to yield improved performance compared to training solely on the target dataset.

TABLE IV
TRAINING AND FINE-TUNING RESULTS ON *VIIRS* AND *MA Landsat* DATASETS USING LANDSAT-MAIN PRETRAINING. THERMAL BAND IS USED FOR TRAINING.

Dataset	Training scheme	U-Net		SegFormer	
	FT: Fine Tune	**Acc**	**mIoU**	**Acc**	**mIoU**
VIIRS	Train	0.79	0.65	0.80	0.66
	Pretrained Inference	0.43	0.22	0.45	0.24
	FT - Frozen encoder	0.77	0.63	0.78	0.63
	FT - Partially Frozen	0.78	0.64	0.80	**0.67**
MA Landsat	Train	0.87	0.77	0.84	0.72
	Pretrained Inference	**0.89**	**0.80**	**0.86**	**0.75**
	FT - Frozen encoder	**0.89**	**0.79**	**0.86**	**0.75**
	FT - Partially Frozen	**0.88**	**0.78**	**0.86**	**0.76**

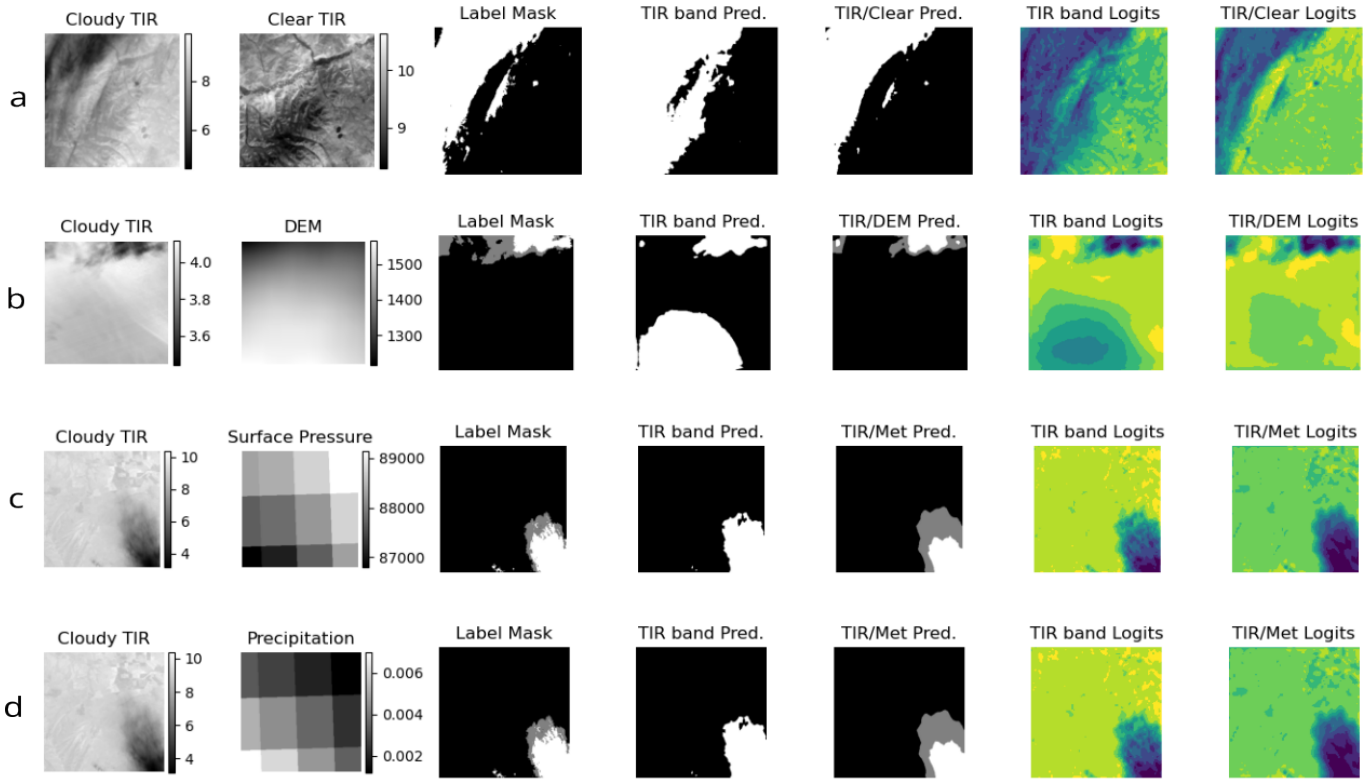

Fig. 5. Examples of inference where auxiliary input contributed to the improvement of accuracy of the model. (a) The model trained only on Cloudy TIR predicts a false positive because the ground in that area exhibited lower than usual thermal values and real clouds were present nearby, causing the region to resemble a cloud. Clear TIR helped resolve this uncertainty. (b) The inclusion of the DEM helps resolve the common infrared based confusion between clouds and high altitude cold terrain (ice/snow), evident in the TIR-only prediction that mistakenly labels part of a mountain as cloud. (c) and (d) The combination of low precipitation and high surface pressure (conditions that suppress deep moist convection and the formation of thick clouds) helps the model more effectively distinguish thin cloud layers from denser cloud cover.

IV. Discussion

TIRAuxCloud addresses the lack of datasets supporting the development of models for continuous, day-and-night cloud detection, which is a critical task for a variety of important remote sensing services. It provides a comprehensive yet computationally practical dataset suite. The dataset combines large-scale automatically labeled subsets with manually annotated data. It also includes auxiliary features that are available during both day and night. These features are related to cloud formation and atmospheric conditions. Together, they complement thermal imaging and help improve cloud detection models. The benchmark results confirm the rich informativeness of the dataset compared to other cloud segmentation studies (e.g. CDNetV2 [17], CloudU-Net [18], Cloud-Net+ [32], SwinCloud [10]).

Our experiments with auxiliary feature combinations showed that attention-equipped models tend to benefit from these additional inputs. However, this improvement remains limited, though it appears consistent across certain feature combinations and architectures. While it is well established that the information contained in auxiliary variables can help reduce model uncertainty and ambiguity, integrating them effectively into a learning framework remains nontrivial. A possible reason for the model's difficulty in exploiting these sources is *unimodal dominance* [43], the tendency of one modality (in this case, the thermal band) to dominate learning. Since the thermal band is closely correlated with the target variable (the cloud mask), the model tends to prioritize this simpler, high-signal pathway while neglecting more complex multimodal relationships. Achieving better performance may require problem-specific architectural solutions. For instance, difference based feature extractors could be employed to explicitly model spectral spatial deviations between cloudy and co-registered clear-sky counterparts. Similarly, coarse resolution meteorological variables could be integrated through a parallel encoder branch operating at lower spatial frequencies.

Experiments on the *MA Landsat* dataset showed that models pretrained on *Landsat Main* outperformed those trained directly on the MA Landsat. Specifically, the pretrained U-Net and SegFormer models outperform those trained on *MA Landsat* by 0.03 in mIoU and by 0.02 in accuracy. This suggests that for cloud segmentation, large-scale automatically labeled data with moderate label noise (e.g., operational Landsat cloud mask products) may be more valuable than smaller manually annotated datasets, indicating that dataset scale can, in some cases, outweigh annotation fidelity.

In contrast, transfer learning to the *VIIRS* dataset requires substantial fine-tuning due to the larger domain gap, yet still achieves comparable or slightly better results than mod-

els trained exclusively on *VIIRS*. More work is needed to strengthen cross-sensor generalization, for example by augmenting the dataset with additional and more diverse sensor sources, leveraging self-supervised and foundation model pretraining, and exploring domain adaptation strategies.

Future work directions:

- Extend dataset coverage across additional sensors, modalities, and geographic regions to enhance generalization.
- Develop models capable of mitigating unimodal dominance to better exploit auxiliary information and improve overall performance.
- Develop self-supervised pretraining strategies for more robust feature learning.
- Explore lightweight architectures optimized for in-orbit deployment.
- Evaluate foundation models on the dataset to assess transferability and cross-domain adaptation.
- Apply explainability techniques to better understand the contribution of auxiliary features.

V. Conclusion

We presented TIRAuxCloud, a novel multi-modal dataset for cloud segmentation. The main contributions include: i) the creation of a large-scale dataset enriched with comprehensive auxiliary data, designed as a testbed for advancing methods in day- and night-time cloud detection; ii) benchmark evaluations using both general-purpose and cloud-specialized models, which establish strong performance baselines and demonstrate the dataset's potential in improving cloud detection methods; and iii) insights into how auxiliary features can contribute to enhanced model accuracy. Further research and experimentation are needed to fully exploit the dataset's potential and ultimately improve the accuracy of critical remote sensing applications operating continuously, day and night.

References

[1] Z. Li, H. Shen, Q. Weng, Y. Zhang, P. Dou, and L. Zhang, "Cloud and cloud shadow detection for optical satellite imagery: Features, algorithms, validation, and prospects," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 188, pp. 89–108, Jun. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271622000934

[2] J. Li, C. Hu, Q. Sheng, J. Xu, C. Zhu, and W. Zhang, "A multi-scale features-based cloud detection method for suomi-npp viirs day and night imagery," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. X-1-2024, pp. 115–122, 2024. [Online]. Available: https://isprs-annals.copernicus.org/articles/X-1-2024/115/2024/

[3] K. Sommer, W. Kabalan, and R. Brunet, "Infrared radiometric image classification and segmentation of cloud structures using a deep-learning framework from ground-based infrared thermal camera observations," *Atmospheric Measurement Techniques*, vol. 18, no. 9, pp. 2083–2101, May 2025, publisher: Copernicus GmbH. [Online]. Available: https://amt.copernicus.org/articles/18/2083/2025/

[4] G. Jedlovec, *Automated Detection of Clouds in Satellite Imagery*. IntechOpen, 10 2009.

[5] S. Mahajan and B. Fataniya, "Cloud detection methodologies: variants and development—a review," *Complex & Intelligent Systems*, vol. 6, no. 2, pp. 251–261, Jul. 2020. [Online]. Available: https://doi.org/10.1007/s40747-019-00128-0

[6] W. Schroeder, I. Csiszar, and J. Morisette, "Quantifying the impact of cloud obscuration on remote sensing of active fires in the Brazilian Amazon," *Remote Sensing of Environment*, vol. 112, no. 2, pp. 456–470, Feb. 2008, publisher: Elsevier BV. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S003442570700199X

[7] I. D. R. Eberhardt, B. Schultz, R. Rizzi, I. D. Sanches, A. R. Formaggio, C. Atzberger, M. P. Mello, M. Immitzer, K. Trabaquini, W. Foschiera, and A. José Barreto Luiz, "Cloud cover assessment for operational crop monitoring systems in tropical areas," *Remote Sensing*, vol. 8, no. 3, 2016. [Online]. Available: https://www.mdpi.com/2072-4292/8/3/219

[8] L. Sun, X. Yang, S. Jia, C. Jia, Q. Wang, X. Liu, J. Wei, and X. Zhou, "Satellite data cloud detection using deep learning supported by hyperspectral data," *International Journal of Remote Sensing*, vol. 41, no. 4, pp. 1349–1371, 2020. [Online]. Available: https://doi.org/10.1080/01431161.2019.1667548

[9] L. Joachim and T. Storch, "CLOUD DETECTION FOR NIGHT-TIME PANCHROMATIC VISIBLE AND NEAR-INFRARED SATELLITE IMAGERY," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-2-2020, pp. 853–860, Aug. 2020, publisher: Copernicus GmbH. [Online]. Available: https://isprs-annals.copernicus.org/articles/V-2-2020/853/2020/

[10] L. Gao, L. Li, J. Yu, X. Zhou, L. Zou, N. Fang, X. Su, and F. Chen, "Swincloud: a hybrid network for cloud detection in thermal infrared remote sensing images," *International Journal of Digital Earth*, vol. 17, no. 1, p. 2409966, 2024. [Online]. Available: https://doi.org/10.1080/17538947.2024.2409966

[11] X. Shi, Y. Fan, L. Sun, X. Liu, C. Liu, and S. Pang, "Cloud detection sample generation algorithm for nighttime satellite imagery based on daytime data and machine learning application," *Scientific Reports*, vol. 14, no. 1, p. 27917, Nov. 2024, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41598-024-78889-z

[12] J. A. Barsi, S. J. Hook, J. R. Schott, N. G. Raqueno, and B. L. Markham, "Landsat-8 thermal infrared sensor (tirs) vicarious radiometric calibration," *Remote Sensing*, vol. 6, no. 11, pp. 11 607–11 626, 2014.

[13] D. K. Hall, G. A. Riggs, N. E. DiGirolamo, and M. O. Román, "Evaluation of modis and viirs cloud-gap-filled snow-cover products for production of an earth science data record," *Hydrology and Earth System Sciences*, vol. 23, no. 12, pp. 5227–5241, 2019. [Online]. Available: https://hess.copernicus.org/articles/23/5227/2019/

[14] S. Foga, P. L. Scaramuzza, S. Guo, Z. Zhu, R. D. Dilley, T. Beckmann, G. L. Schmidt, J. L. Dwyer, M. Joseph Hughes, and B. Laue, "Cloud detection algorithm comparison and validation for operational landsat data products," *Remote Sensing of Environment*, vol. 194, pp. 379–390, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425717301293

[15] Q. Li, W. Lu, and J. Yang, "A hybrid thresholding algorithm for cloud detection on ground-based color images," *Journal of Atmospheric and Oceanic Technology*, vol. 28, no. 10, pp. 1286 – 1296, 2011. [Online]. Available: https://journals.ametsoc.org/view/journals/atot/28/10/jtech-d-11-00009_1.xml

[16] S. Zhang, H. Zhang, B. Zhang, Z. Cui, and C. Xiao, "An Improved Fmask Algorithm for Cloud Detection Applied to Hyperspectral Satellite," *Acta Optica Sinica*, vol. 43, no. 24, p. 2428009, Dec. 2023. [Online]. Available: https://www.researching.cn/articles/OJ44da8052fb3529e8

[17] J. Guo, J. Yang, H. Yue, H. Tan, C. Hou, and K. Li, "Cdnetv2: Cnn-based cloud detection for remote sensing imagery with cloud-snow coexistence," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 700–713, 2021.

[18] C. Shi, Y. Zhou, B. Qiu, D. Guo, and M. Li, "Cloudu-net: A deep convolutional neural network architecture for daytime and nighttime cloud images' segmentation," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 10, pp. 1688–1692, 2021.

[19] J. Li, T. Xue, J. Zhao, J. Ge, Y. Min, W. Su, and K. Zhan, "High-resolution cloud detection network," 2024. [Online]. Available: https://arxiv.org/abs/2407.07365

[20] X. Wu, Z. Shi, and Z. Zou, "A geographic information-driven method and a new large scale dataset for remote sensing cloud/snow detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 174, pp. 87–104, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271621000290

[21] J. M. Haynes, Y.-J. Noh, S. D. Miller, K. D. Haynes, I. Ebert-Uphoff, and A. Heidinger, "Low cloud detection in multilayer scenes using satellite imagery with machine learning methods," *Journal of Atmospheric and Oceanic Technology*, vol. 39, no. 3, pp. 319–334, 2022. [Online]. Available: https://doi.org/10.1175/JTECH-D-21-0084.1

[22] P. Scaramuzza, "Landsat 8 collection 1 cloud truth mask validation set," 2021, u.S. Geological Survey data release, Accessed 2025-08-01. [Online]. Available: https://doi.org/10.5066/P9U71ZVU

[23] ——, "Landsat 8 collection 2 cloud truth mask validation set," 2021, accessed 2025-08-01. [Online]. Available: https://doi.org/10.5066/P9FI4A0Y

[24] World Bank, "Light every night (len): daily viirs-dnb and dmsp-ols nighttime light data, 1992–2020," 2021, developed in collaboration with NOAA and University of Michigan. [Online]. Available: https://worldbank.github.io/OpenNightLights/

[25] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, B. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J.-R. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, F. Félix, E. Gérard, Z. Görlich, A. Jönsson, P. Källberg, T.-Y. Koh, M. Matricardi, A. McNorton, I. Moye, S. Pöttgen, D. Röthlisberger, M.-L. Sicre, E. Oña, R. van der Schalie, T. van Noije, S. van der Veen, N. Wedi, and E. Zsóter, "The era5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.

[26] C. Wang, S. Platnick, K. Meyer, Z. Zhang, and Y. Zhou, "A machine-learning-based cloud detection and thermodynamic-phase classification algorithm using passive spectral observations," *Atmospheric Measurement Techniques*, vol. 13, no. 5, pp. 2257–2277, May 2020. [Online]. Available: https://amt.copernicus.org/articles/13/2257/2020/

[27] C. H. White, A. K. Heidinger, and S. A. Ackerman, "Evaluation of Visible Infrared Imaging Radiometer Suite (VIIRS) neural network cloud detection against current operational cloud masks," *Atmospheric Measurement Techniques*, vol. 14, no. 5, pp. 3371–3394, May 2021. [Online]. Available: https://amt.copernicus.org/articles/14/3371/2021/

[28] H. Zhou, C.-H. Kao, C. P. Phoo, U. Mall, B. Hariharan, and K. Bala, "Allclear: A comprehensive dataset and benchmark for cloud removal in satellite imagery," in *Advances in Neural Information Processing Systems*, A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, Eds., vol. 37. Curran Associates, Inc., 2024, pp. 53 571–53 597. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2024/file/60095e1d7ebb292dbba93c4d8f7b2463-Paper-Datasets_and_Benchmarks_Track.pdf

[29] L. Fabio, D. Piga, M. Umberto, and E. G. Safouane, "Benchcloudvision: A benchmark analysis of deep learning approaches for cloud detection and segmentation in remote sensing imagery," 2024. [Online]. Available: https://arxiv.org/abs/2402.13918

[30] X. Zou, S. Zhang, K. Li, S. Wang, J. Xing, L. Jin, C. Lang, and P. Tao, "Adapting vision foundation models for robust cloud segmentation in remote sensing images," 2024. [Online]. Available: https://arxiv.org/abs/2411.13127

[31] S. Mohajerani and P. Saeedi, "Cloud-net: An end-to-end cloud detection algorithm for landsat 8 imagery," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2019, pp. 1029–1032.

[32] S. Mohajerani and P. Saeedi, "Cloud-Net+: A Cloud Segmentation CNN for Landsat 8 Remote Sensing Imagery Optimized with Filtered Jaccard Loss Function," vol. 2001.08768, 2020.

[33] C. Aybar, L. Ysuhuaylas, J. Loja, K. Gonzales, F. Herrera, L. Bautista, R. Yali, A. Flores, L. Diaz, N. Cuenca, W. Espinoza, F. Prudencio, V. Llactayo, D. Montero, M. Sudmanns, D. Tiede, G. Mateo-García, and L. Gómez-Chova, "CloudSEN12, a global dataset for semantic understanding of cloud and cloud shadow in Sentinel-2," *Scientific Data*, vol. 9, no. 1, p. 782, Dec. 2022, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41597-022-01878-2

[34] C. Aybar, L. Bautista, D. Montero, J. Contreras, D. Ayala, F. Prudencio, J. Loja, L. Ysuhuaylas, F. Herrera, K. Gonzales, J. Valladares, L. A. Flores, E. Mamani, M. Quiñonez, R. Fajardo, W. Espinoza, A. Limas, R. Yali, A. Alcántara, M. Leyva, R. Loayza-Muro, B. Willems, G. Mateo-García, and L. Gómez-Chova, "Cloudsen12+: The largest dataset of expert-labeled pixels for cloud and cloud shadow detection in sentinel-2," *Data in Brief*, vol. 56, p. 110852, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352340924008163

[35] Earth Resources Observation and Science (EROS) Center, "Landsat 8–9 operational land imager / thermal infrared sensor level-1, collection 2 [dataset]," 2020, accessed via USGS/EROS M2M API. [Online]. Available: https://doi.org/10.5066/P975CC9B

[36] NOAA National Geophysical Data Center, "2-minute gridded global relief data (etopo2) v2," 2006. [Online]. Available: https://doi.org/10.7289/V5J1012Q

[37] Copernicus Global Land Service, "Land cover 2019 (raster 100 m), global, yearly – version 3," 2019. [Online]. Available: https://land.copernicus.eu/en/products/global-dynamic-land-cover/copernicus-global-land-service-land-cover-100m-collection-3-epoch-2019-globe

[38] J. Muñoz Sabater, "Era5-land hourly data from 1950 to present," 2019, accessed on 2025-08-01. [Online]. Available: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land

[39] H. E. Beck, N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood, "Present and future Köppen-geiger climate classification maps at 1-km resolution," *Scientific Data*, vol. 5, 2018. [Online]. Available: https://doi.org/10.1038/sdata.2018.214

[40] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2015, pp. 234–241.

[41] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," in *Neural Information Processing Systems (NeurIPS)*, 2021.

[42] S. Kondylatos, N. I. Bountos, I. Prapas, A. Zavras, G. Camps-Valls, and I. Papoutsis, "Probabilistic machine learning for noisy labels in earth observation," *Scientific Reports*, vol. 15, no. 1, p. 35890, 2025.

[43] X. Zheng, Y. Lyu, L. Jiang, D. P. Paudel, L. V. Gool, and X. Hu, "Reducing Unimodal Bias in Multi-Modal Semantic Segmentation with Multi-Scale Functional Entropy Regularization."