

# Geospatial Active Learning for Efficient Data Annotation: A case study on cool roof detection

YOGENDRA KUMAR, Plaksha University, India

ANUPAM SOBTI, Plaksha University, India

Geospatial data is foundational to understanding systemic changes that our society is undergoing. Whether it is changing land cover, changes in public infrastructure, or natural resources, supervised machine learning algorithms are being found increasingly useful. However, annotation of geospatial data is difficult due to large areas of interest and geographic variation in different areas. In this work, we use the example of cool roof classification to demonstrate an active learning method that is specifically designed to take benefit from the geospatial setting. Cool roofs are reflective coatings on roofs that keep the building cooler and thus also reduce electricity requirements for cooling. In prior work[16], pipelines for this classification have been developed through building footprint extraction and reflectance estimation through feature-based classification. We adopt a similar methodology for classification and in this paper, we propose two geospatial-specific AL methods that benefit from temporal uncertainty estimate and overconfident predictions resampling. We verify the efficacy of our data annotation by verifying the classification results through an actual decrease in land surface temperature. We also develop a tool for fast and easy labelling of annotations with active learning in the loop. This helps us reduce the annotation time for a city from a few days to a few hours while achieving similar accuracies in ~50% training iterations. Our thermal validation using Landsat-derived Land Surface Temperature (LST) data confirms that annotated cool roofs consistently exhibit lower surface temperatures during summer months, with statistically significant differences supported by the Mann-Whitney U Test.

CCS Concepts: • Computing methodologies → Active learning settings;

Additional Key Words and Phrases: Active Learning, Cool Roofs, ADS, CTPV

## ACM Reference Format:

Yogendra Kumar and Anupam Sobi. 2025. Geospatial Active Learning for Efficient Data Annotation: A case study on cool roof detection. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS '25), July 22–25, 2025, Toronto, ON, Canada*. ACM, New York, NY, USA, 22 pages. <https://doi.org/10.1145/3715335.3735492>

## 1 Introduction

The intensifying effect of Urban Heat Islands (UHI) presents a significant challenge to the sustainability of urban development, particularly in rapidly growing cities across India. As urban areas become increasingly dense, urban heat retention caused by the extensive use of heat-absorbing materials, such as concrete and asphalt, results in elevated temperatures that are often several degrees higher than surrounding rural regions. This temperature differential not only exacerbates energy consumption, particularly during the hot summer months but also increases the demand for air conditioning, which in turn contributes to higher greenhouse gas emissions and worsens global warming [26].

---

Authors' Contact Information: **Yogendra Kumar**, yogendra.kumar@plaksha.edu.in, Plaksha University, Mohali, Punjab, India; **Anupam Sobi**, anupam.sobti@plaksha.edu.in, Plaksha University, Mohali, Punjab, India.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

Among the various strategies proposed to mitigate the effects of UHI, cool roofs have gained prominence as a cost-effective and environmentally sustainable solution. These roofs, made from materials that reflect more sunlight and emit heat efficiently, have been shown to reduce both surface and indoor temperatures. Reflecting a significant portion of the incoming solar radiation, cool roofs help to lower the ambient temperature of urban areas, thus alleviating the strain on air conditioning systems and reducing energy consumption. Furthermore, the reduction in urban temperatures also contributes to better public health and improved urban comfort [1, 17]. Multiple studies have indicated that cool roofs can achieve temperature reductions of up to 3 °C in urban environments [2], providing a substantial contribution to mitigating the impact of urban heat.

However, despite the considerable benefits of cool roofs, several challenges hinder their large-scale implementation, particularly in Indian cities. Roof materials can vary widely even within the same city and urban environments are characterized by complex patterns of building density, roof geometry, and surrounding microclimates. These factors pose significant hurdles in accurately identifying and evaluating cool roofs using traditional methods such as manual surveys or small-scale inspections [18]. Remote sensing technologies, particularly high-resolution satellite imagery, offer a promising alternative by enabling large-scale monitoring of urban thermal conditions. However, even with advances in remote sensing, two critical challenges remain: (1) the high cost and inefficiency of manual annotation required to train machine learning models and (2) the difficulty of calibration due to the inherent variability of urban landscapes, such as differences in roof materials, seasonal effects, and low resolutions sensor inconsistencies. These factors make it difficult to apply classification models effectively in diverse geographic regions, leading to inaccuracies in cool roof identification [28].

To address these challenges, in this paper, we implemented a hybrid approach that combines satellite-based observations with a Multilayer Perceptron (MLP) to efficiently classify cool roofs in large urban areas. **Active learning** has proven to be a valuable tool in remote sensing applications, allowing significant reductions in both annotation effort and calibration inconsistencies—i.e., variations in model performance due to domain shift (changes in geographic regions), data quality, or environmental conditions that affect the consistency of predictions across space and time—while simultaneously improving model performance [24]. Our methodology incorporates an active learning framework designed to minimize annotation costs while improving model performance. We propose two new acquisition functions, named Cross Time Prediction Variance (CTPV) and Adaptive Disparity Acquisition Sampling (ADS), to strategically select the most informative samples for labelling. Active learning reduces the need for extensive manual annotation while enhancing classification accuracy. Furthermore, our approach addresses temporal inconsistencies using CTPV and overconfident predictions using ADS by prioritizing the most informative and confident samples, ensuring more reliable predictions across varying climatic and urban conditions.

In addition to the active learning framework, this study also performs a detailed thermal analysis of cool roofs using Landsat-8 thermal data. Our analysis demonstrates that cool roofs consistently show a reduction in LST compared to non-cool roofs. This analysis highlights the long-term cooling benefits of cool roofs, which are particularly pronounced during the summer months when the effects of UHI are most severe [3]. To evaluate the localized cooling effects of cool roofs, we compare the LST of cool roofs with their neighbouring buildings within a 100-meter radius.

To streamline the process of labelling large datasets and facilitate the scaling of cool-roof identification, we develop a user-friendly annotation platform. This platform is designed to simplify the labelling of satellite imagery and provide a scalable solution for the creation of high-quality data sets. It is adaptable to other remote sensing applications and offers broad utility in various geospatial research fields.

The main contributions of this work are:

- **Novel Active Learning Acquisition Functions:** We propose Cross-Time Prediction Variance (CTPV) and Adaptive Disparity Sampling (ADS) for geospatial active learning in cool roof classification. CTPV prioritizes temporally stable samples, ADS targets overconfident misclassifications.
- **Thermal Analysis of Cool Roofs:** Using Landsat-8 thermal data, we demonstrate that cool roofs consistently lower land surface temperature (LST) compared to non-cool roofs. This indicates correct classifications and the efficiency of cool roofs in reducing temperature.
- **Satellite Data Annotation Tool:** A versatile and user-friendly platform for labelling satellite data, facilitating large-scale identification of cool roofs and adaptable to other remote sensing tasks.

The remainder of this paper is organized as follows. Section 2 reviews the existing methods in remote sensing-based cool roof identification and active learning frameworks. Section 3 details our data sources, feature extraction methods, dimensionality reduction techniques and the MLP-based classification approach. Section 4 details the experimental setup for the acquisition functions and the associated training procedure. Section 5 presents experimental results, comparing the performance of different active learning strategies in urban environments. Finally, Section 6 summarizes the main findings, discusses limitations, and suggests future research directions.

## 2 Literature Review

Urban Heat Islands (UHI) pose a significant environmental challenge in urbanized areas, where infrastructure materials such as asphalt and concrete absorb and retain heat, leading to elevated local temperatures and increased energy demand [26]. These increases in temperature contribute to health risks, higher cooling costs, and amplified climatic impacts, necessitating effective mitigation strategies.

An approach widely studied is the implementation of **cool roofs**, which utilize materials with high solar reflectance and thermal emissivity to reduce heat absorption [1, 17]. Empirical evidence suggests that increasing roof reflectivity can lower ambient temperatures by up to 3°C in urban settings [2]. This cooling effect improves indoor thermal comfort and contributes to overall energy efficiency.

### 2.1 Use of Geospatial Data in Urban Analysis

Geospatial data is a widely applicable tool in urban analysis. Recent studies have explored the use of multi-spectral satellite data, such as Sentinel-2 imagery, for environmental and land-use monitoring applications [13, 19]. The detection of brick kilns using spectral bands, as demonstrated by Imaduddin et al. [13], exemplifies how remote sensing techniques can identify industrial activities that contribute to urban heat and air pollution. Similarly, land use and land cover (LULC) classification methods utilizing machine learning in Google Earth Engine have provided accurate urban mapping results [19].

Remote sensing has become an essential tool for examining UHIs and Identifying cool roofs, using satellite platforms (for example, Sentinel-2 and Landsat-8) to measure surface reflectance, albedo, and land surface temperature (LST) [14, 16, 28]. High-resolution satellite imagery enables both spatial and temporal analysis of urban thermal conditions. However, issues such as cloud cover, atmospheric interference, and coarse sensor resolution can restrict data quality [18]. Automated detection and mapping of cool roofs remain key challenges due to urban variability and lack of labelled datasets [7]. Recent advances in **remote sensing and machine learning**, particularly **Active Learning (AL)**, are effective in geospatial applications. While active learning remains underexplored for cool-roof classification. To address

these gaps, we critically examine the limitations of traditional AL approaches and propose novel acquisition strategies tailored for cool-roof detection.

## 2.2 Limitations of Existing Active Learning Strategies

Conventional active learning (AL) has been widely applied to remote sensing and image classification tasks, often using *uncertainty-based sampling* to prioritize high-uncertainty samples and *diversity-based sampling* to ensure feature space coverage [18]. However, in the context of cool-roof classification, these methods often fall short due to domain-specific challenges like temporal variation and material heterogeneity.

Temporal Inconsistency-Based Active Learning (TIR-AL)[27] and Temporal Output Discrepancy (TOD)[12] address challenges related to model forgetting by selecting samples based on inconsistencies during various training phases, while TOD assesses sample informativeness by evaluating prediction discrepancies across different optimization steps. Despite proving effective in multiclass classification and semantic segmentation, these strategies encounter limitations in binary classification tasks like cool-roof detection. Specifically, TOD-based methods may struggle to identify subtle differences between roof reflectance, and clustering approaches such as TypiClust [10] often require a large number of samples to construct useful decision boundaries, thereby diminishing sample efficiency.

To address these limitations, we introduce two novel acquisition functions—**Cross Time Prediction Variance (CTPV)** and **Adaptive Disparity Acquisition Sampling (ADS)**—that are specifically designed to maximize label efficiency for cool-roof classification:

- **CTPV** selects a minimal subset of samples with high temporal prediction variance, capturing cases most affected by seasonal changes and improving generalization across time.
- **ADS** emphasizes label efficiency by selecting only those samples that lie in high-disparity regions of the feature space, helping the model quickly learn class boundaries with fewer examples.

Together, CTPV and ADS reduce the number of labelled samples needed to reach high classification accuracy, without compromising performance. By explicitly accounting for urban-specific challenges, our method achieves greater sample efficiency than existing AL techniques, making it well-suited for scalable deployment in resource-constrained urban monitoring scenarios.

## 2.3 Classification of Roof Materials

Classifying roof materials using satellite and aerial imagery is a crucial step in understanding urban heat dynamics and identifying cool roofs. Traditional classification approaches rely on supervised learning models that require large annotated datasets. However, these methods face several challenges, including spectral similarity between materials, seasonal variations, and variations in roof conditions.

Several recent studies have explored automated rooftop classification methods. Park et al. [20] proposed a CNN-based remote sensing approach to detect and classify rooftops in urban areas to support cool roof applications. Their study demonstrated how aerial imagery and machine learning techniques could be used to classify roofs based on surface reflectance. Similarly, Chen and Li [6] developed a deep learning-based rooftop detection method using convolutional neural networks (CNN) and Mask R-CNN for post-earthquake building assessment. Their approach successfully extracted rooftop features from aerial images, highlighting the potential of deep learning models in building classification tasks.

A recent study, *What's Up On The Roof: Tracking Cool Roofs in India with Satellite Imaging* [16], introduced a cool roof tracking system that classifies roofs into cool and non-cool categories based on their reflectance properties. Their Manuscript submitted to ACM

work demonstrated the feasibility of using satellite-derived reflectance. Similarly to [16], we utilize satellite imagery to classify cool and non-cool roofs. However, such methods are dependent on high-quality city aerial data and require calibration. Our approach expands on their methodology by integrating active learning to reduce the annotation cost. We verify the correctness of cool roofs by showing the land surface temperature (LST) differences between cool and non-cool roofs. We also show the effect of cool roofs on LST by comparing cool roofs and their neighbouring structures within a radius of 100m to account for localized microclimatic effects.

Annotating high-resolution satellite imagery is both resource-intensive and time-consuming. Existing annotation tools often lack flexibility or require extensive customization [5, 29]. To overcome these issues and reduce the cost and time of annotation, we developed a user-friendly **annotation tool** that is adapted to the classification needs of the cool roof.

*Key Contributions of This Work:* Building on existing research on active learning, remote sensing, and UHI mitigation, our study presents novel methodologies and tools that address persistent gaps in these domains. The primary contributions are as follows:

- (1) We propose **Cross Time Prediction Variance (CTPV)**, an acquisition function that accounts for temporal variations in satellite imagery. Unlike simple random-date approaches, CTPV identifies dates where models exhibit inconsistent predictions and leverages these insights to drive sample selection. Comprehensive comparisons indicate that CTPV-selected samples deliver more robust model performance than randomly chosen ones, highlighting the reliability and stability of the method.
- (2) We introduce **Adaptive Disparity Acquisition Sampling (ADS)**, a novel approach that combines uncertainty-based selection with a disparity metric to pinpoint highly informative samples. Empirical evaluations demonstrate that ADS surpasses established methods e.g., random selection, entropy-based acquisition [22], margin sampling [4], and BALD [11] in both data efficiency and classification accuracy, significantly lowering annotation costs in active learning pipelines.
- (3) We also elucidate the *thermal* effect of cool roofs, consistently observing lower LST values compared to non-cool roofs. The statistical significance of these differences is supported by the Mann-Whitney U test, confirming the robustness of our observations.
- (4) Lastly, we present a **satellite data annotation tool** designed to accelerate the labelling process for high-resolution roof reflectance datasets. Its intuitive interface and adaptable scripts reduce technical barriers to large-scale data annotation, supporting a wide range of geospatial research and operational tasks.

In general, these contributions advance the methodology in active learning and remote sensing for roof classification and calibration, offering practical insights for urban climate resilience planning and strengthening the role of cool roofs in combating UHIs.

### 3 Methodology

This section outlines the framework and methodologies employed to address the research questions posed earlier. We first describe the acquisition and preprocessing of Sentinel-2, followed by the architecture of our MLP classifier and its training procedure. Subsequently, we introduce our active learning framework, including two complementary acquisition functions (CTPV and Adaptive Disparity Sampling). Finally, we detail our thermal analysis approach using Landsat data to compare cool and non-cool roofs.

### 3.1 Data Acquisition and Preprocessing

*Satellite Imagery and Building Polygons.* Following Lalwani et al. [16], we leveraged Sentinel-2 satellite imagery to extract 8 spectral bands (e.g., Aerosol, Blue, Green, Red, NIR) for feature generation. These bands offer comprehensive spectral information that is highly indicative of the material and reflectance properties [8]. Building footprints were obtained from the Google Open Buildings dataset [23], which provides detailed polygonal outlines of rooftops across diverse geographic regions.

*Spatial Alignment.* Next, each roof polygon from the Open Buildings dataset was spatially aligned with the corresponding Sentinel-2 image tiles. This step allowed us to precisely locate the spectral information for each roof polygon.

*Feature Extraction and Normalization.* From each roof polygon, we extracted the pixel values for eight Sentinel-2 bands. These were then aggregated into a feature vector, ensuring that we captured both spectral and spatial properties. All features were scaled into the range [0, 1] using a MinMaxScaler to prevent any single band or feature from dominating during model training.

*Dimensionality Reduction (PCA).* Although we focus on eight spectral bands, additional derived indices or texture characteristics can lead to high-dimensional representations. To balance expressiveness and computational efficiency, we employed Principal Component Analysis (PCA) and retained 30 principal components. These components captured most of the variance and served as input to our downstream classification model.

### 3.2 Model Architecture

We adopt a **multilayer perceptron (MLP)** for the binary classification task of identifying cool versus non-cool roofs. The primary motivation for selecting an MLP is its computational efficiency and previous findings by Lalwani et al.[16], which demonstrated the effectiveness of a compact MLP on roof classification tasks. After dimensionality reduction (Section 3.1) yields 30 principal components, these components feed into the MLP architecture described below.

*Network Design.* As summarized in Table 1, the MLP comprises:

- **Input Layer:** Accepts the 30-dimensional PCA feature vector.
- **Hidden Layers:** Two fully connected (dense) layers with 64 and 8 neurons, respectively. We employ the Rectified Linear Unit (ReLU) activation function in each hidden layer to introduce nonlinearity.
- **Output Layer:** A single neuron with a sigmoid activation function produces a probability score in [0, 1], indicating the likelihood of a roof being classified as cool.

*Parameter Summary.* Table 1 details the dimensions and total parameters for each layer. Implementation specifics, optimization details, and hyperparameter choices are provided in Section 4.2, along with our training procedure.

### 3.3 Active Learning Framework

Although supervised ML has been widely successful, the task of labelling extensive imagery in varied urban environments is costly and labour-intensive. To address this issue, we incorporate active learning (AL), which systematically chooses the most *informative* samples for labelling, as shown in Figure 1, thus reducing labelling expenses while improving model efficiency.

Table 1. Summary of the MLP architecture for cool-roof classification.

Layer	Output Shape	Trainable Params
Input (PCA Features)	(None, 30)	-
Dense (Hidden 1)	(None, 64)	$30 \times 64 + 64$
Dense (Hidden 2)	(None, 8)	$64 \times 8 + 8$
Output (Sigmoid)	(None, 1)	$8 \times 1 + 1$

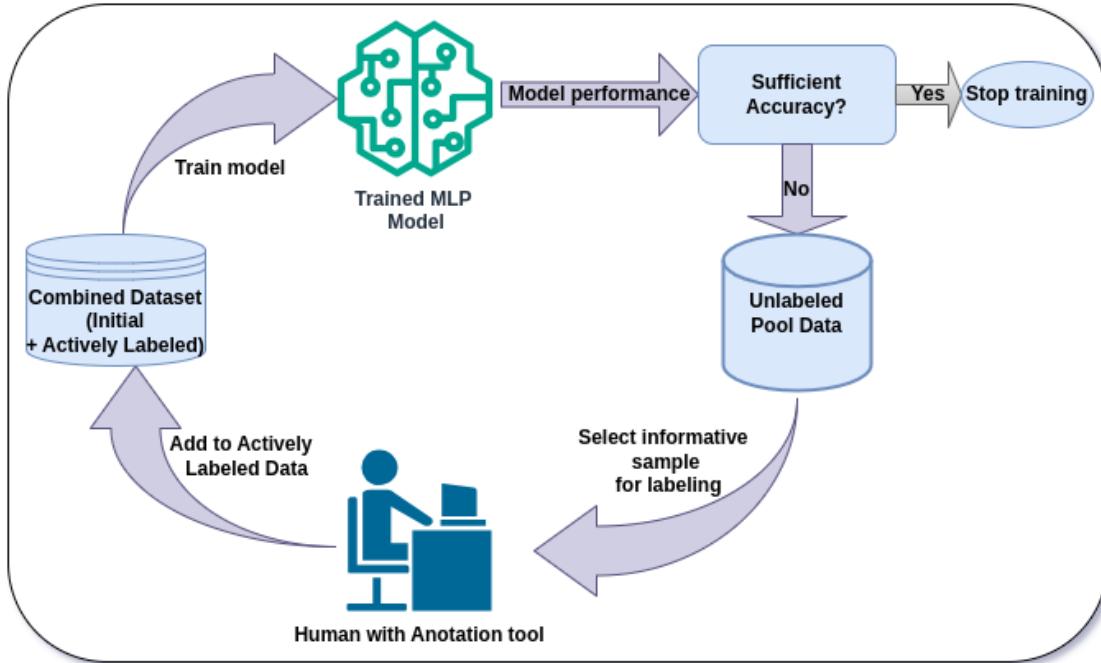


Fig. 1. Active Learning Workflow for Training a Model: The iterative process of selecting informative samples from an unlabeled pool, annotating them, and retraining the model until performance meets the desired threshold.

In our framework as shown in figure 2, two complementary strategies are employed: Cross-Time Prediction Variance (CTPV) and Adaptive Disparity Sampling (ADS), both designed to select high-impact samples for labelling.

CTPV aims to identify samples from time points where the model predictions are more stable. By considering the variation in model predictions across multiple time points, CTPV helps prioritize those periods where the model exhibits consistency, thereby ensuring temporal robustness.

ADS, on the other hand, focuses on selecting the best samples, where the model confidence is highest. This strategy not only targets confident samples but also enforces diversity to prevent overfitting to highly similar instances, ensuring the model learns from a broader range of data.

### 3.3.1 Cross-Time Prediction Variance (CTPV).

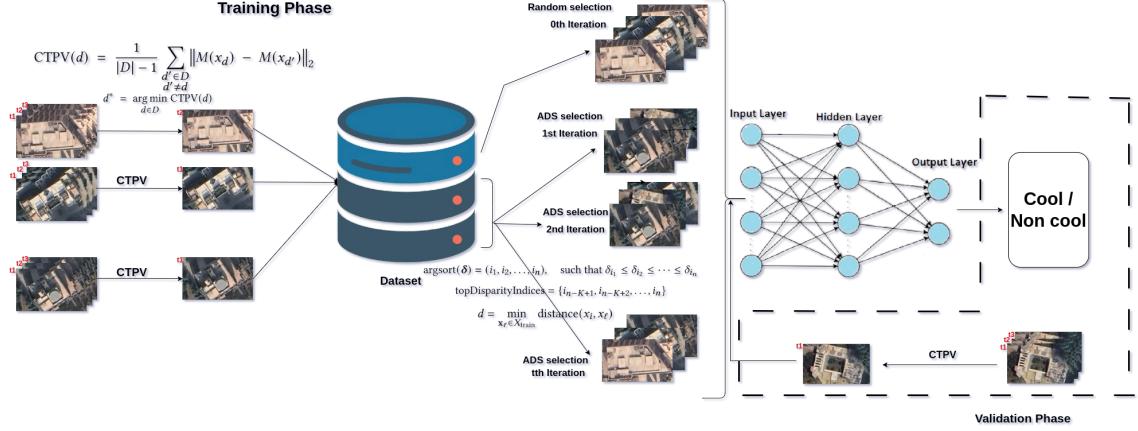


Fig. 2. Acquisition Function Workflow for Model Training: CTPV selects the most informative samples across different time steps, while ADS selects samples from the pool dataset. The model is then retrained until its performance meets the desired threshold.

*Definition.* Let  $D = \{d_1, d_2, \dots, d_n\}$  be a collection of dates or time points on which Sentinel-2 imagery is available. For each date  $d \in D$ , we form a feature vector  $x_d \in \mathbb{R}^m$ . A trained model  $M(\cdot)$  produces a prediction  $M(x_d)$  in  $\mathbb{R}^p$ . We define the Cross-Time Prediction Variance for date  $d$  as:

$$\text{CTPV}(d) = \frac{1}{|D|-1} \sum_{\substack{d' \in D \\ d' \neq d}} \|M(x_d) - M(x_{d'})\|_2, \quad (1)$$

where  $\|\mathbf{v}\|_2$  denotes the Euclidean norm in  $\mathbb{R}^p$ .

*Acquisition Criterion.* CTPV selects the date

$$d^* = \arg \min_{d \in D} \text{CTPV}(d), \quad (2)$$

implicitly favouring time points whose predictions are the most consistent (mean pairwise distance) relative to other dates. In an active learning loop, this *stability-driven* criterion can reduce label redundancy and guide annotation to temporally robust samples. Algorithm 1 outlines the procedure to calculate CTPV on all available dates and select the one with the minimum score. The algorithm evaluates the model predictions  $M(x_d)$  for each date  $d$ , and returns both the selected date  $d^*$  and its corresponding features  $x_{d^*}$ .

### 3.3.2 Adaptive Disparity Sampling (ADS).

*Definition.* Let  $X_{\text{pool}} = \{x_1, x_2, \dots, x_N\}$  be a collection of unlabelled samples, and let  $X_{\text{train}} = \{x_1, x_2, \dots, x_L\}$  represent the labelled training set. Given a trained model  $M(\cdot)$ , the model produces a predicted probability  $p_i = M(x_i) \in [0, 1]$  for each sample  $x_i \in X_{\text{pool}}$ . For each sample, we compute a *disparity score*  $\delta_i$  to quantify the certainty of the model's prediction:

$$\delta_i = |p_i - 0.5|, \quad (3)$$

**Algorithm 1** CTPV Acquisition Algorithm

---

**Require:** A trained model  $M$ , a set of dates  $D$ , and feature vectors  $\{x_d \mid d \in D\}$ .  
**Ensure:**  $(d^*, x_{d^*}, \{\text{CTPV}(d) \mid d \in D\})$ .

```

1: Initialize an empty dictionary PREDICTIONS.
2: for each  $d \in D$  do
3:    $\hat{y}_d \leftarrow M(x_d)$                                      ▷ Model prediction for date  $d$ 
4:   PREDICTIONS[d]  $\leftarrow \hat{y}_d$ 
5: end for
6: Initialize CTPVSCORES[d]  $\leftarrow 0$  for all  $d \in D$ .
7: for each  $d \in D$  do
8:   for each  $d' \in D$  with  $d' \neq d$  do
9:     CTPVSCORES[d]  $\leftarrow \text{CTPVSCORES}[d] + \|\text{PREDICTIONS}[d] - \text{PREDICTIONS}[d']\|_2$ 
10:    end for
11:   CTPVSCORES[d]  $\leftarrow \text{CTPVSCORES}[d]/(|D| - 1)$ 
12: end for
13:  $d^* \leftarrow \arg \min_{d \in D} \text{CTPVSCORES}[d]$ 
14: return  $(d^*, x_{d^*}, \{\text{CTPVSCORES}[d] : d \in D\})$ 
```

---

where a low  $\delta_i$  indicates higher uncertainty and suggests that the sample lies closer to the decision boundary. However, we observed that overconfident misclassifications were equally problematic as high-uncertainty predictions due to confusing inputs.

*Acquisition Criterion.* To select the most informative samples, we prioritize those with high disparity scores, which are the samples with low uncertainty. Let  $K = \min(5 \cdot n, N)$  be the number of top samples to consider, where  $n$  is the number of instances to acquire. We select the top- $K$  samples with the highest disparity scores:

$$\begin{aligned} \text{argsort}(\delta) &= (i_1, i_2, \dots, i_n), \quad \text{such that } \delta_{i_1} \leq \delta_{i_2} \leq \dots \leq \delta_{i_n} \\ \text{topDisparityIndices} &= \{i_{n-K+1}, i_{n-K+2}, \dots, i_n\} \end{aligned}$$

and then apply a diversity criterion to ensure that the selected samples are diverse concerning the labelled data. For each candidate sample  $x_i \in \text{topDisparityIndices}$ , we compute the minimum distance  $d$  to the labeled data:

$$d = \min_{x_\ell \in X_{\text{train}}} \text{distance}(x_i, x_\ell), \tag{4}$$

and enforce a diversity threshold  $\text{divThreshold}$  (e.g., the 25th percentile of the disparity scores). Only samples with a distance  $d$  greater than the threshold are selected. If fewer than  $n$  samples meet the diversity criterion, we fill the remaining quota with the next most confident samples.

Algorithm 2 outlines the procedure to compute the disparity scores, select the most confident samples, enforce the diversity criterion, and optionally fill in the remaining samples through a fallback mechanism.

### 3.4 Thermal Analysis of Cool Roofs

We conducted a thermal analysis using remote sensing data to validate the quality and correctness of the labels produced through our annotation platform. Specifically, we aimed to verify whether the annotated cool roofs indeed exhibited lower land surface temperatures, thus reinforcing the credibility of our data labelling pipeline.

**Algorithm 2** Adaptive Disparity Sampling

---

**Require:** Trained model  $M(\cdot)$ ; Unlabeled pool  $\mathbf{X}_{\text{pool}} \in \mathbb{R}^{N \times m}$ ; Labeled training set  $\mathbf{X}_{\text{train}} \in \mathbb{R}^{L \times m}$ ; Number of instances to acquire,  $n$ .

**Ensure:** Indices of the  $n$  selected samples from  $\mathbf{X}_{\text{pool}}$ .

```

1: Step 1: Compute Disparity Scores
2: Evaluate  $M(\mathbf{X}_{\text{pool}})$  to obtain predicted probabilities  $\mathbf{p} \in \mathbb{R}^N$ 
3: for  $i = 1$  to  $N$  do
4:    $\delta_i \leftarrow |p_i - 0.5|$ 
5: end for

6: Step 2: Select High-Disparity Candidates
7:  $K \leftarrow \min\{5 \cdot n, N\}$ 
8:  $\text{topDisparityIndices} \leftarrow \text{argsort}(\delta)[-K:]$  ▷ Indices of top- $K$  disparity scores

9: Step 3: Enforce Diversity
10:  $\text{divThreshold} \leftarrow \text{percentile}(\delta, 25)$ 
11:  $\text{selected} \leftarrow \{\}$ 
12: for  $i$  in  $\text{topDisparityIndices}$  (descending order of  $\delta_i$ ) do
13:    $\mathbf{x}_i \leftarrow \mathbf{X}_{\text{pool}}[i]$ 
14:    $d \leftarrow \min_{\mathbf{x}_\ell \in \mathbf{X}_{\text{train}}} \text{distance}(\mathbf{x}_i, \mathbf{x}_\ell)$ 
15:   if  $d > \text{divThreshold}$  then
16:      $\text{selected} \leftarrow \text{selected} \cup \{i\}$ 
17:   end if
18:   if  $|\text{selected}| \geq n$  then
19:     break
20:   end if
21: end for

22: Step 4: Fallback (If Needed)
23: if  $|\text{selected}| < n$  then
24:    $\text{remainingNeeded} \leftarrow n - |\text{selected}|$ 
25:    $\text{fallback} \leftarrow \text{topDisparityIndices} \setminus \text{selected}$ 
26:    $\text{fallback} \leftarrow \text{fallback}[0 : \text{remainingNeeded}]$ 
27:    $\text{selected} \leftarrow \text{selected} \cup \text{fallback}$ 
28: end if
29: return  $\text{selected}$ 
```

---

We used the USGS Earth Explorer to acquire the Landsat 8-9 OLI/TIRS Collection 2 Level 2 dataset [25] for measuring Land Surface Temperature (*LST*) at the city scale. This dataset was processed to extract *LST* values to evaluate the thermal performance of annotated cool roofs in comparison to non-cool roofs.

To assess seasonal performance differences and validate spatial labels, our analysis included:

- Identifying the predicted locations of cool roofs across the city and extracting their corresponding *LST* values in degrees Celsius.
- Analyzed seasonal variations in *LST* for both cool and non-cool roofs across all months, supported by the Mann-Whitney U Test to assess the statistical significance of temperature differences.
- Comparing the *LST* of cool roofs with their neighbouring buildings within a 100-meter radius to understand localized microclimatic cooling effects.

The thermal effect assessment, based on comparisons of **mean LST values** between cool and non-cool roofs, revealed a consistent reduction in surface temperatures for the annotated cool roofs, particularly during summer months. These

Manuscript submitted to ACM

findings confirm that the cool roof labels generated by our annotation tool are thermally meaningful and correspond to real-world cooling benefits.

### 3.5 Satellite Data Annotation Tool

To facilitate efficient and accurate labelling of satellite imagery, we develop an interactive **annotation dashboard** designed to streamline the manual labelling process. The dashboard integrates Google Maps imagery with Sentinel-2 satellite images of the target area, using latitude and longitude to guide human annotators. Annotators visually assess these image pairs to verify whether the identified buildings contain cool roofs.

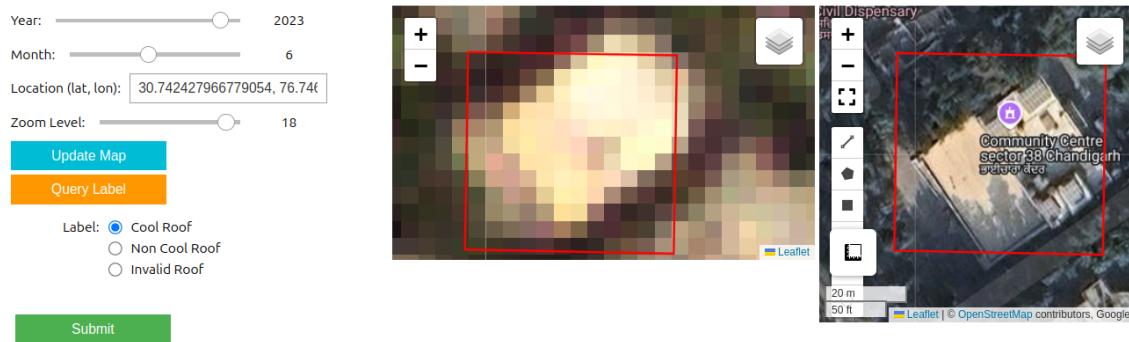


Fig. 3. The annotation dashboard interface displaying Google Maps[9] and Sentinel-2 RGB Bands imagery [8] with roof squares bounding box to polygon mapping for guided labelling.

The annotation workflow follows an active learning approach to optimize labelling efficiency. Initially, a small manually labeled data set is used to train a baseline model. The model then generates predictions for a larger, unlabeled dataset. To further refine the data set, an *acquisition function* classifies samples based on their informativeness or confidence, prioritizing them for manual review. This iterative process significantly reduces the annotation effort while maintaining high accuracy.

The dashboard provides several key functionalities:

- **Interactive satellite image visualization:** Enables precise manual annotation of roofs using intuitive selection tools.
- **Support for multiple annotation formats:** Ensures compatibility with standard machine learning workflows for downstream tasks.
- **Adaptability for broader remote sensing applications:** Allows for easy modification of scripts and configurations to accommodate different annotation tasks.

By integrating these features, our annotation tool improves the efficiency of dataset creation, supporting high-quality training data generation for machine learning applications in remote sensing.

## 4 Experimental Setup and Evaluation Metrics

We conducted experiments to assess the effectiveness of the CTPV acquisition function compared to random sampling using a cool roof dataset from Chandigarh City. Additionally, we evaluated the performance of our ADS acquisition function in comparison with various acquisition functions documented in the literature. The evaluation metric used

across all experiments was test accuracy, measured on a held-out validation set. This allowed us to quantify the impact of each acquisition function in selecting informative samples and improving model generalization in the context of cool roof classification.

#### 4.1 Datasets

Using our satellite data annotation tool, we have labelled the cool roofs dataset of buildings located in the city of Chandigarh. This dataset includes 457 cool roofs and 451 non-cool roofs. Statistically, our data sets are evenly distributed, as depicted in Fig. 4. We used Sentinel-2 data from three distinct dates: 4 June, 9 June and 16 June, incorporating eight bands for pre-processing and feature extraction.

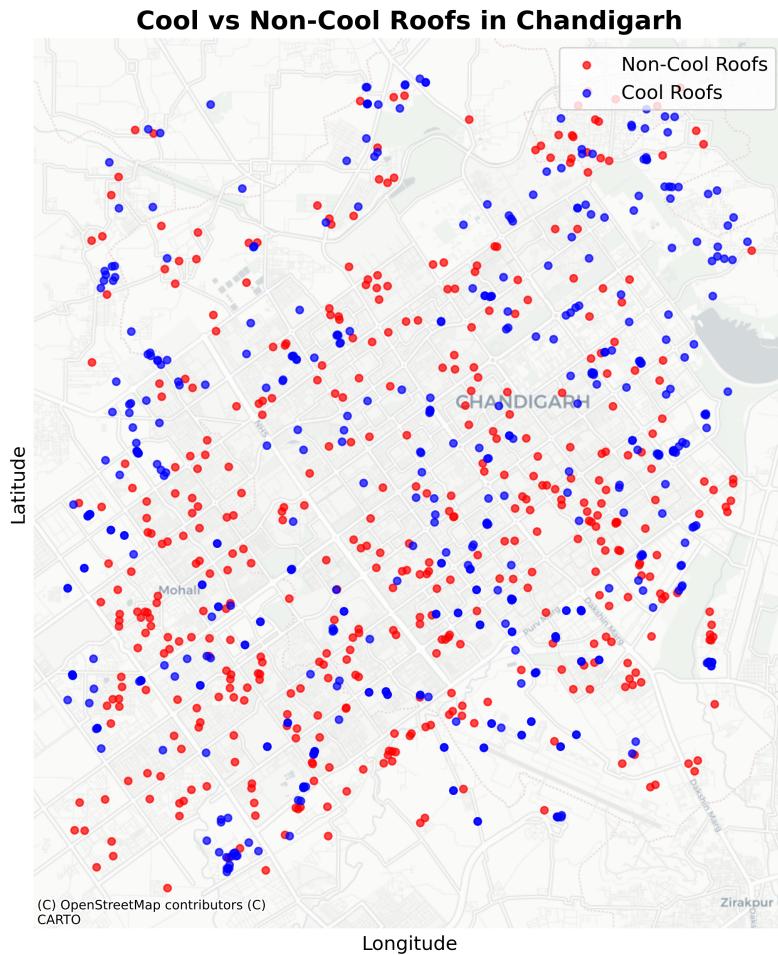


Fig. 4. Spatial distribution of 457 annotated cool roofs and 451 non-cool roofs in Chandigarh city. Blue markers represent cool roofs, while red markers represent non-cool roofs.

## 4.2 Training Procedure and Hyperparameters

The experiments employ an MLP as described in Section 3.2, implemented using PyTorch 2.3.1[21]. Initially, 98 samples are randomly selected to train the initial model and 112 samples are used as a fixed test set. Each active learning iteration, as discussed in Sections 4.3 and 4.4, begins with retraining the model on newly labelled samples while keeping the hyperparameters unchanged. We incorporate 20 samples per iteration into the active learning cycle, ensuring a balanced class distribution. The Adam Optimizer with an initial learning rate of  $10^{-3}$  is used for optimization and training is limited to 30 epochs for each iteration. All experiments were carried out using an NVIDIA RTX A5000 GPU on a server.

## 4.3 Comparison of CTPV Acquisition Function Performance

We conducted a comparative study to assess the effectiveness of the proposed CTPV acquisition function against a baseline random sampling method. The main objective was to determine if CTPV could select more informative samples by utilizing temporal prediction variance, thus enhancing the model's test accuracy with fewer labelled examples.

The experiment was structured to emulate a pool-based active learning environment. In each iteration, the acquisition function chose a subset of samples from the unlabelled pool to be labelled. These newly labelled samples were later used to update the model incrementally. The comparison included the following acquisition functions:

- **CTPV (proposed):** Utilizes temporal prediction variance to select the most informative samples.
- **Baseline (random sampling):** Selects samples randomly without considering the uncertainty or informativeness of the model.

We evaluate the effectiveness of the acquisition functions by measuring test accuracy over several iterations. The results of this experiment are discussed in the Results section 5, where we demonstrate the effectiveness of CTPV in outperforming the baseline and other acquisition strategies.

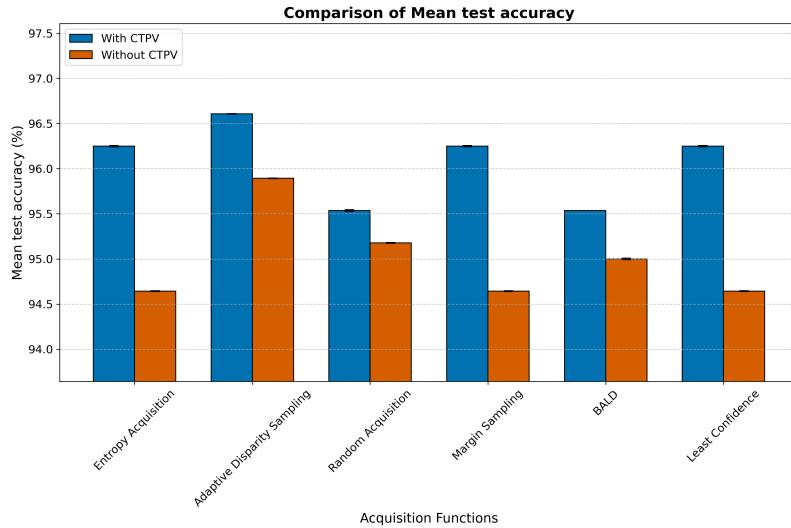


Fig. 5. Performance comparison of the CTPV acquisition function and the baseline random sampling strategy based on test accuracy.

#### 4.4 Comparison of ADS Acquisition Function Performance

In addition to CTPV, we introduce a new active learning acquisition strategy called Adaptive Disparity Sampling (ADS). The effectiveness of ADS was tested against a range of established baseline strategies, such as Entropy Acquisition, Random Sampling, Least Confidence, Margin Sampling, and BALD (Bayesian Active Learning by Disagreement).

The ADS approach utilizes differences in predictions across multiple samples from the pool data to identify highly informative data points. By focusing on input space areas with significant uncertainty or lacking representation, ADS seeks to enhance model generalization while reducing the number of labelled samples needed.

We evaluate the effectiveness of the acquisition functions by measuring test accuracy over several iterations. To ensure consistency, the experiments were repeated using different random seeds to address variability. The average performance of these iterations was documented and standard deviations were calculated to signify the robustness of each method. Findings are detailed in the Results section 5, underscoring the benefits of ADS in achieving higher test accuracy with fewer labelled samples than standard methods.

#### 4.5 Summary of the Workflow

The suggested workflow utilizes a structured strategy to assess active learning acquisition functions for efficient sample selection in machine learning tasks. This approach seeks to minimize labelling efforts while sustaining or improving model performance, especially in tasks dealing with temporal and spectral data.

- (1) **Data Acquisition and Preprocessing:** Collect Sentinel-2 satellite imagery and extract building footprints across multiple temporal instances. Pre-process data to ensure consistency and accuracy.
- (2) **Feature Engineering:** Extract raw spectral bands and derive additional characteristics relevant to urban thermal analysis and reflective roof classification.
- (3) **Dimensionality Reduction:** Utilize Principal Component Analysis (PCA) to reduce the dimensionality of spectral and spatial features, retaining the most informative components.
- (4) **MLP Model Training and Validation:** Train and validate a Multi-Layer Perceptron (MLP) model on labelled data to classify cool and non-cool roofs. The model was initially trained on 98 labelled samples and evaluated on a fixed test set of 112 samples.
- (5) **Active Learning Workflow:** Implement an iterative active learning loop where the most informative unlabelled roof samples are selected using acquisition functions. In each iteration, 20 samples are selected from the pool dataset and used to retrain the model. This process aims to minimize labelling costs while maintaining high model performance.

### 5 Results and Discussion

This section examines and interprets the outcomes of our proposed methodology when applied to the Chandigarh cool roof data set. We examine how well the multilayer perceptron classifier (MLP) identifies cool roofs, assess the efficiency of the active learning framework in minimizing annotation workload while preserving high classification accuracy, and offer an in-depth evaluation of land surface temperature (LST) measurements. Lastly, we present a broader discussion of our findings, highlighting significant insights and possible limitations.

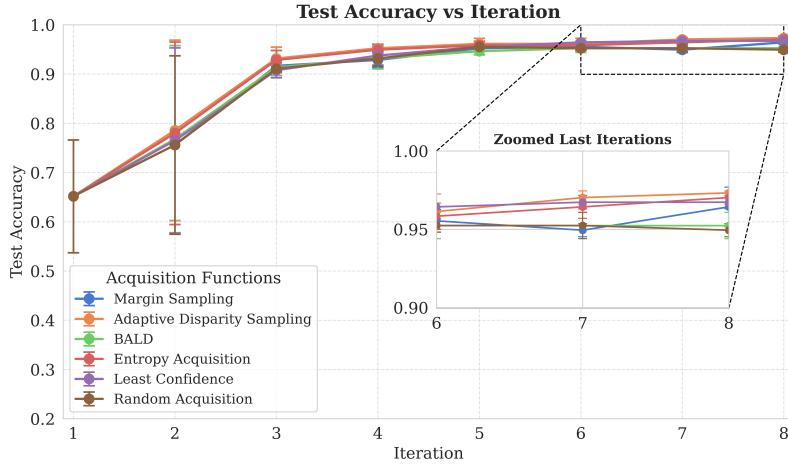


Fig. 6. Test accuracy performance of the proposed Adaptive Disparity Acquisition Sampling (ADS) strategy compared to baseline strategies, including Cross-Time Prediction Variance (CTPV), over multiple active learning iterations. Each iteration adds 20 samples from the pool dataset. Error bars indicate the standard deviation of accuracy across five independent runs. While performance differences are subtle at higher sample counts, the proposed ADS approach shows greater improvements at lower sample sizes, demonstrating its utility in low-data regimes.

### 5.1 Cool Roof Classification Performance

The classification performance of our MLP model in Chandigarh demonstrates strong results. The model achieved an accuracy of 97.02%, for distinguishing between cool and noncool roofs. This indicates a high level of reliability in the model's predictions,

### 5.2 Active Learning Efficacy

**5.2.1 Reduction in Annotation Effort.** Figure 6 presents the learning curves comparing the conventional labelling approach (random sampling) and other acquisition functions versus our active learning (AL) strategy. Notable observations include:

- **Steeper Performance Gain:** Active Learning (AL) methods such as ADS achieved over 96% test accuracy with just 210 labelled samples, whereas random sampling failed to reach 96% accuracy even after annotating 300 samples. This highlights the efficiency of AL strategies in reducing the overall annotation workload while achieving higher model performance.
- **Faster Convergence:** The model trained with AL converged to high precision within fewer iterations, indicating an efficient selection of ambiguous or diverse samples.

We explored both confidence-based and diversity-based acquisition functions. Adaptive disparity dynamically monitoring changes in confidence scores, this strategy prioritized samples in which the model exhibited fluctuating confidence between training epochs. The approach further reduced redundancy in labelled examples.

The confidence intervals in Figure 6 are computed based on standard deviations derived from five independent experimental runs with the same train-test split to maintain consistency. Specifically, each acquisition strategy (e.g.,

Table 2. Comparison of different active learning acquisition strategies based on test performance metrics of final iteration. The reported mean and standard deviation values are calculated over five independent runs. While differences are small across methods, the proposed ADS approach consistently achieves the highest mean accuracy.

Method	Mean Test Accuracy	Std Test Accuracy	Mean Test Loss	Std Test Loss
Margin Sampling	0.9494	0.0042	0.1557	0.0165
Adaptive Disparity Sampling	0.9702	0.0042	0.1381	0.0147
BALD	0.9524	0.0084	0.1522	0.0150
Entropy Acquisition	0.9643	0.0073	0.1348	0.0142
Least Confidence	0.9673	0.0042	0.1455	0.0176
Random Acquisition	0.9524	0.0084	0.1596	0.0201

ADS, CTPV) is evaluated across five independent runs. The standard deviations reported in Table 2 are calculated using this same procedure to ensure robustness and reliability.

Our method outperformed baseline random sampling, indicating that an informed sample selection strategy substantially reduces annotation costs without sacrificing accuracy. Although we acknowledge that differences between active learning methods are marginal at higher iteration numbers. However, improvements are more significant during early-stage training, where fewer samples are available. This observation is particularly relevant for urban policy applications where reducing annotation costs while maintaining accuracy is critical.

Using this method, we analyzed 213,000 buildings and identified 4,315 cool roofs. We then validated 457 cool roofs using our dashboard to examine their Land Surface Temperature (LST) characteristics and assess the real-world thermal effect of these classified cool roofs. The subsequent subsection quantifies the thermal performance comparison between these detected cool roofs and non-cool roofs.

### 5.3 Real-World Thermal Effect of Cool Roofs

The results highlight the importance of cool roof technologies as an effective sustainable approach to improving urban thermal comfort and increasing energy efficiency, especially in areas experiencing high summer temperatures. Future research may aim to include more detailed spatial data, extend the duration of datasets, and consider socioeconomic elements to more accurately assess the wide-ranging effects of implementing cool roofs in urban settings.

**5.3.1 Land Surface Temperature Differences.** Using the thermal band, we computed LST maps for both *cool* and *non-cool* rooftops. Figure 7 presents a histogram of the LST values at the building level in Chandigarh.

- **Overall Temperature Reduction:** During the summer months, cool roofs exhibited a lower Land Surface Temperature (LST) by up to 2 °C compared to their non-cool counterparts during peak daytime heating.

To ensure the robustness of our temperature comparison, we performed a Mann-Whitney U test for each month of the year. The results are presented in Table 4. Statistically significant differences were observed between cool and non-cool roofs for most months ( $p$ -value < 0.05). Exceptions include August and November, where the differences were not statistically significant, likely due to monsoon cloud cover or seasonal effects. This analysis confirms that cool roofs generally exhibit lower temperatures compared to their non-cool counterparts throughout the year.

These results support previous research indicating that reflective roof technologies can substantially reduce daytime surface temperatures in cities[15]. Nevertheless, the precise extent of these cooling effects is influenced by local morphological elements like the density of buildings, nearby vegetation, and the city's structural layout.

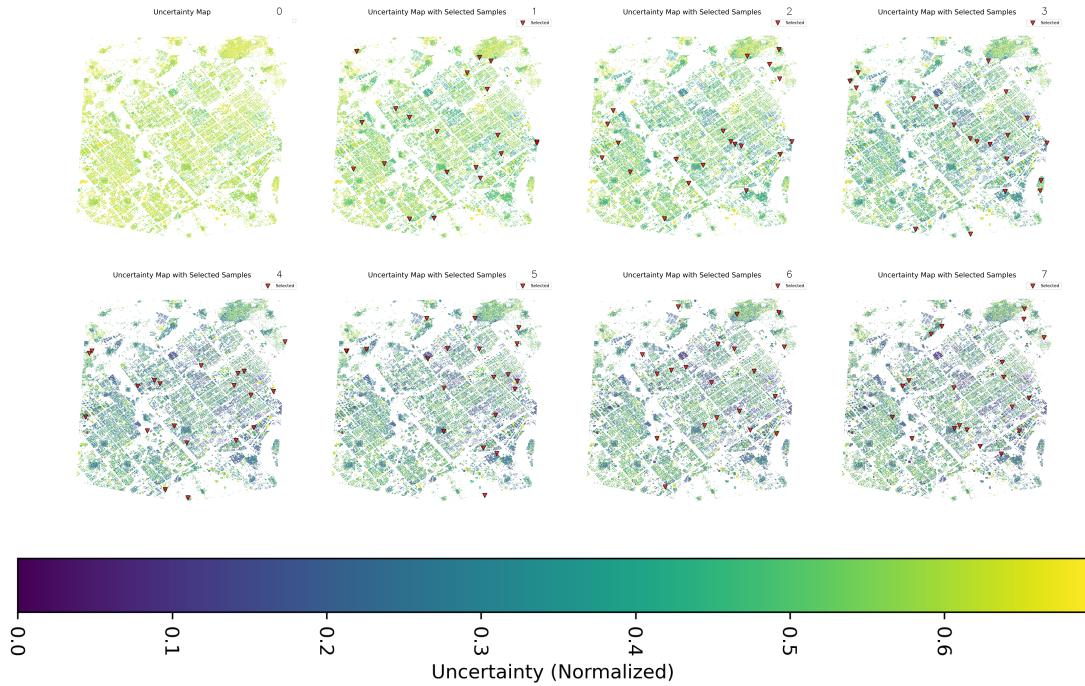


Table 3. Visualization of geographically distributed sample selection, the uncertainty map, and the iterative selection process using the Adaptive Disparity Acquisition Sampling (ADS) approach. Each subfigure represents a distinct uncertainty map at different iterations, showcasing the progressive refinement of model performance.

Observation Period	U-Statistic	P-Value	Statistically Significant
January 2024	117659.0	0.00021878	Yes
February 2024	90325.0	0.00127635	Yes
March 2024	91038.0	0.00235955	Yes
April 2024	94712.5	0.03478603	Yes
May 2024	91637.0	0.00386227	Yes
June 2024	93089.5	0.01168246	Yes
July 2024	88060.5	0.00014803	Yes
August 2024	104562.5	0.70262783	No
September 2024	83874.0	1.21e-06	Yes
October 2024	94580.0	0.03200287	Yes
November 2024	97708.5	0.17618136	No
December 2024	111666.0	0.02929046	Yes

Table 4. Mann-Whitney U Test results for the statistical significance of land surface temperature differences between cool roofs and their non-cool roof neighbours. Significant differences ( $p\text{-value} < 0.05$ ) are observed for most months, except during the monsoon period (August) and the transition month of November.

5.3.2 *Land Surface Temperature Differences Between Cool Roofs and Neighbouring Buildings.* This study examines the effect of cool roofs on land surface temperature (LST) by comparing the temperatures of buildings with cool roofs with

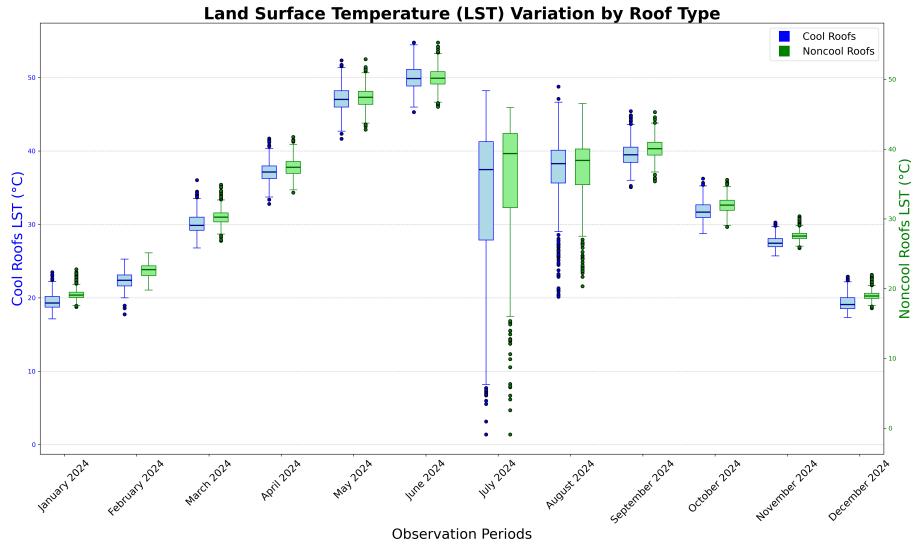


Fig. 7. Monthly Land Surface Temperature (LST) analysis in Chandigarh showcasing seasonal temperature variations between cool and non-cool roofs. The plot highlights differences across all months, emphasizing the effectiveness of cool roofs in reducing LST, particularly during the hotter months of April, May, and June.

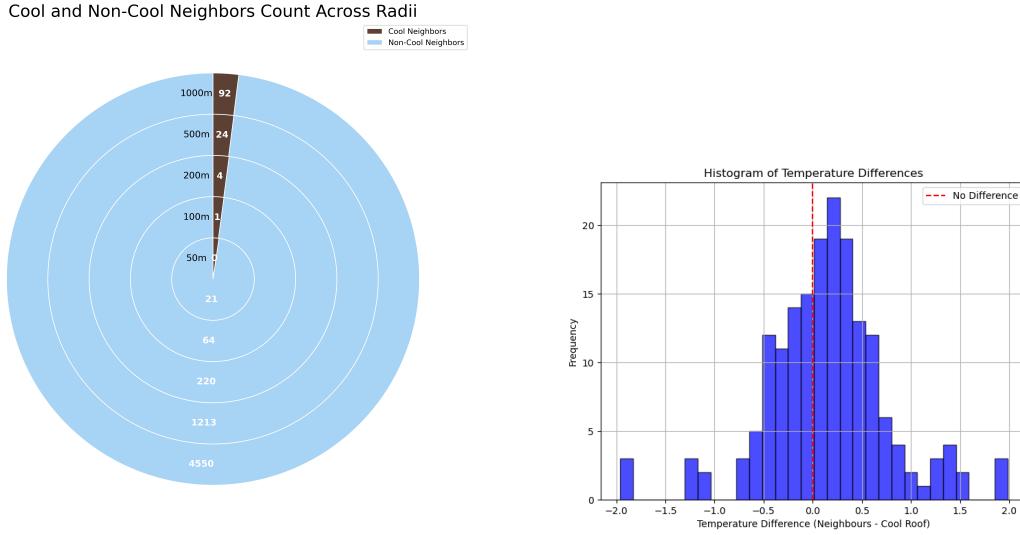
those of their neighbouring structures within a 100-meter radius. The analysis demonstrates that cool roof buildings exhibit significantly lower surface temperatures than their immediate surroundings.

Figure 8a illustrates the distribution of the average number of cool and non-cool roofs within varying radii from a target cool roof. We observe that a radius of 100m provides a suitable compromise between sample size and minimizing variance from unrelated temperature influences, such as vegetation or water bodies.

Figure 8b presents a histogram that illustrates the distribution of temperature differences between the cool-roof buildings and their neighbouring structures. The right-skewed distribution, with most data points showing a positive temperature difference. This trend suggests that cool-roof buildings generally maintain lower surface temperatures than their surroundings.

However, in some cases, an inverse pattern is observed, as depicted in Figure 8b. This deviation occurs primarily due to the presence of dense vegetation around certain roofs or a high concentration of cool roofs within the 100-meter radius. Figure 9 visually highlights these conditions, demonstrating that external environmental factors can sometimes mitigate the cooling advantage of individual cool roofs.

In summary, our model consistently achieves more than 97% accuracy in distinguishing cool versus non-cool roofs, while our active learning framework significantly reduces the annotation cost. These findings underscore the potential of reflective roofing to reduce urban heat, with observed average LST reductions of 2–3°C in Chandigarh. Future expansions will explore higher-resolution thermal data and additional climate regions to further validate and refine our methodology.



(a) Comparison of the average number of cool and non-cool roofs among neighbors across different radii (50m, 100m, 200m).

(b) Histogram of average LST differences between cool roof buildings and their neighbors within a 100-meter radius.

Fig. 8. Neighborhood-level analysis of cool roofs: (a) The plot highlights the count of cool roofs in proximity and the effectiveness of a 100m radius in providing meaningful thermal contrast. and (b) temperature differences highlighting localized cooling effects. The right-skewed distribution indicates that cool roofs generally maintain lower temperatures compared to their surroundings.

## 6 Conclusion and Future Work

In this study, we have investigated the cooling effects of reflective cool roofs, which present a promising and cost-effective solution to mitigate the Urban Heat Island (UHI) effect, especially in rapidly growing cities like those in India. We developed a hybrid approach that leverages remote sensing data along with machine learning techniques, particularly a Multilayer Perceptron (MLP) classifier, to identify cool roofs across large urban landscapes. Our approach also integrates active learning (AL), which not only improves classification performance but also reduces the time and effort involved in annotating large satellite datasets.

One of the key contributions of this work is the introduction of two novel active learning acquisition functions: **Cross Time Prediction Variance (CTPV)** and **Adaptive Disparity Acquisition Sampling (ADS)**. These strategies have proven to be highly effective in selecting the most informative samples for annotation, which ultimately leads to a more accurate model with fewer labelled data points. Our results demonstrate that both CTPV and ADS outperform a range of state-of-the-art active learning showing significant gains in both classification accuracy and data efficiency.

Our thermal analysis, based on Landsat-8 thermal data, further supports the benefits of cool roofs and the efficiency of our pipeline. We observed a significant reduction in Land Surface Temperature (LST), particularly in the summer months, highlighting their effectiveness in providing thermal relief during peak urban heat stress conditions. This finding underscores the importance of cool roofs in improving urban thermal comfort and contributing to energy savings, particularly in tropical regions such as India.



Fig. 9. Illustration of an area where the expected cooling effect of cool roofs is not observed due to surrounding vegetation and a high density of cool roofs within the 100-meter radius.

The study also introduced a lightweight, user-friendly annotation tool to accelerate satellite image labelling, enabling scalable cool roof detection across cities. This tool bridges the gap between manual labelling and automated pipelines, facilitating the rapid deployment of annotated datasets in other geospatial research contexts.

Our findings underscore the dual value of cool roofs: first, as a proven passive technique to lower urban surface temperatures, and second, as a use case for scalable, efficient annotation frameworks in satellite image analysis. Active learning, in particular, emerges as a powerful enabler, allowing rapid model adaptation with minimal manual effort. In particular, we demonstrate Geo AI methods such as CTPV and ADS for the binary classification task of cool building detection and validate our methodology through LST-based thermal verification.

While our results are promising, several limitations warrant future investigation. The current analysis is based on imagery from a limited temporal window, which does not account for seasonal or diurnal variability in the classification of roofs. Extending the temporal span of observations will allow a more comprehensive understanding of the year-round cool roofs. Similarly, the 100-meter resolution of Landsat-8 LST data imposes spatial constraints, particularly for densely built areas where a single building may occupy only a fraction of a pixel. Future work should consider higher-resolution sources or UAV-based thermal imaging to improve precision.

Although active learning reduces the need for large labelled datasets, a minimal level of expert oversight remains essential. Exploring semi-automated or crowdsourced annotation methods may further streamline this process and

enable rapid dataset expansion. Incorporating metrics like precision, recall, and F1-score in future evaluations will offer a more complete view of model performance, especially in low-data or generalization settings.

Several future directions are worth pursuing. Our current implementation of CTPV selects a single observation per building from a limited number of temporal snapshots. Expanding this to include more dates and diverse weather conditions may improve both model robustness and annotation reliability. Furthermore, integrating LST data with urban features such as building density, vegetation cover, and wind dynamics could offer a more holistic view of urban heat behaviour. Another promising extension is to investigate the influence zones of cool roofs by aggregating across multiple neighborhoods. Analyzing larger contiguous clusters of cool roofs could provide insights into their broader collective cooling effects on urban microclimates, particularly in dense urban settings. Adding socioeconomic layers to our analysis may also help identify priority regions for cool roof interventions, especially in underserved and vulnerable communities. In addition, investigating the combined effect of cool roofs with other urban heat mitigation strategies (e.g., green roofs, reflective pavements, and urban forestry) can inform the design of integrated climate resilience plans. Lastly, there is significant scope for policy engagement. Evaluating the role of incentives and government programs in scaling cool roof adoption will be crucial, particularly in regions where urban climate resilience is both urgent and underfunded.

In conclusion, this work advances the state of the art in remote sensing and machine learning for climate adaptation. It provides practical tools and scalable frameworks for identifying cool roofs efficiently, thus contributing to sustainable urban planning in the context of rising global temperatures and expanding urban footprints.

### Acknowledgments

We would like to express our gratitude to Indorama Ventures Center for Clean Energy for their invaluable support and contribution to this research. Their assistance has been instrumental in its completion.

### References

- [1] Hashem Akbari, Sarah Bretz, Dan Kurn, and Jay Hanford. 2001. Peak power and cooling energy savings of high-albedo roofs. *Energy and Buildings* 25, 2 (2001), 117–126.
- [2] Hashem Akbari, Surabi Menon, and Arthur Rosenfeld. 2008. Surfaces that cool urban heat islands: Global warming and mitigating urban heat islands. *Building and Environment* 43, 4 (2008), 601–605.
- [3] Maria F. Astudillo, Karin Lundgren, Dan Chen, and Vojtěch Máca. 2019. Quantifying urban heat island intensity and its association with land use and surface parameters in a mid-size European city. *Journal of Environmental Management* 241 (2019), 541–553.
- [4] Maria-Florina Balcan, Andrei Broder, and Tong Zhang. 2007. Margin Based Active Learning. In *Learning Theory*, Nader H. Bshouty and Claudio Gentile (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 35–50.
- [5] Mariana Belgiu and Lucian Drăguț. 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* 114 (2016), 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- [6] Mengge Chen and Jonathan Li. 2019. Deep convolutional neural network application on rooftop detection for aerial image. *arXiv preprint arXiv:1910.13509* (2019).
- [7] Hok S. Cheung, Wenjing Lou, and Vivian W. Y. Tam. 2019. Cool roof performance and the maintenance of their solar reflectance. *Renewable and Sustainable Energy Reviews* 102 (2019), 122–133.
- [8] Copernicus Sentinel-2. 2023. MSI Level-2A BOA Reflectance Product. [https://doi.org/10.5270/S2\\_-6eb6imz](https://doi.org/10.5270/S2_-6eb6imz) Collection 0.
- [9] Google. 2024-2025. Google Maps. <https://www.google.com/maps>. Accessed multiple times between 2024-06-03 and 2025-01-15.
- [10] Guy Hacohen, Avihu Dekel, and Daphna Weinshall. 2022. Active Learning on a Budget: Opposite Strategies Suit High and Low Budgets. *arXiv preprint arXiv:2202.02794* (2022).
- [11] Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. 2011. Bayesian Active Learning for Classification and Preference Learning. *arXiv:1112.5745* [stat.ML] <https://arxiv.org/abs/1112.5745>
- [12] Siyu Huang, Tianyang Wang, Haoyi Xiong, Bihan Wen, Jun Huan, and Dejing Dou. 2022. Temporal Output Discrepancy for Loss Estimation-Based Active Learning. *IEEE Transactions on Neural Networks and Learning Systems* 34 (2022), 1–15.

- [13] Syed Imaduddin, Yusuf Ahmed Khan, Khushboo Mirza, and B. K. Bhadra. 2023. Detection of Brick Kilns Using Multi-Spectral Bands of Sentinel-2 Imagery. In *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*. 496–503. <https://doi.org/10.1109/AISC56616.2023.10085085>
- [14] Bahareh Kalantar, Shattri Mansor, Biswajeet Pradhan, Koen Renner, Price Ashburn, and Fahmi Abir. 2020. Urban surface albedo and its influence on city-scale surface temperature: A case study of major cities in the United States. *ISPRS Journal of Photogrammetry and Remote Sensing* 159 (2020), 62–78.
- [15] Samiran Khorat, Debashish Das, Rupali Khatun, Sk Mohammad Aziz, Prashant Anand, Ansar Khan, Mattheos Santamouris, and Dev Niyogi. 2024. Cool roof strategies for urban thermal resilience to extreme heatwaves in tropical cities. *Energy and Buildings* 302 (2024), 113751. <https://doi.org/10.1016/j.enbuild.2023.113751>
- [16] Varchita Lalwani, Anupam Sotbi, and Vishal Garg. 2024. What's Up On The Roof: Tracking Cool Roofs in India with Satellite Imaging. *ACM J. Comput. Sustain. Soc.* 2, 4, Article 43 (Nov. 2024), 26 pages. <https://doi.org/10.1145/3685696>
- [17] Ronnen Levinson, Hashem Akbari, and Paul Berdahl. 2010. Measuring solar reflectance—Part I: Defining a metric to characterize roof reflectance. *Solar Energy* 84, 9 (2010), 1717–1744.
- [18] Xiang Li, Lorenzo Bruzzone, and Antonio Plaza. 2019. Active learning for hyperspectral image classification with scarce training samples. *IEEE Geoscience and Remote Sensing Letters* 16, 2 (2019), 264–268.
- [19] Arpittha M, S. A. Ahmed, and Harishnaika N. 2023. Land use and land cover classification using machine learning algorithms in Google Earth Engine. *Earth Science Informatics* 16 (2023), 3057–3073. <https://doi.org/10.1007/s12145-023-01073-w>
- [20] Jaehyeong Park, Sangun Park, and Juyoung Kang. 2024. Detecting and classifying rooftops with a CNN-based remote-sensing method for urban area cool roof application. *Energy Reports* 11 (2024), 2516–2525.
- [21] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems* 32 (2019).
- [22] C. E. Shannon. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27, 3 (1948), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- [23] Wojciech Sirko, Sergii Kashubin, Marvin Ritter, Abigail Annkah, Yasser Salah Eddine Bouchareb, Yann Dauphin, Daniel Keysers, Maxim Neumann, Moustapha Cisse, and John Quinn. 2021. Continental-Scale Building Detection from High Resolution Satellite Imagery. arXiv:2107.12283 [cs.CV] <https://arxiv.org/abs/2107.12283>
- [24] Devis Tuia, Frédéric Ratle, Fabio Pacifici, Mikhail F. Kanevski, and William J. Emery. 2009. Active Learning Methods for Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing* 47, 7 (2009), 2218–2232. <https://doi.org/10.1109/TGRS.2008.2010404>
- [25] U.S. Geological Survey. 2023. Landsat 8-9 OLI/TIRS Collection 2 Level 2 Science Products. U.S. Geological Survey data release. <https://doi.org/10.5066/P9OGBGM6>
- [26] James A. Voogt and Timothy R. Oke. 2004. Urban heat islands: hotter nights, increased energy use, and climate change. *Encyclopedia of Energy* 5 (2004), 517–528.
- [27] Tianjiao Wan, Yutao Dou, Kele Xu, Zijian Gao, Bo Ding, Dawei Feng, and Huaimin Wang. 2024. Temporal Inconsistency-Based Active Learning. *ICASSP 2024 - IEEE International Conference on Acoustics, Speech and Signal Processing* (2024).
- [28] Rong Wang, Zhen Li, Lin Chen, and Xiaoli Zhang. 2021. Remote sensing of urban heat islands: an overview of methods, challenges, and perspectives. *Remote Sensing* 13, 12 (2021), 2357.
- [29] L. Yang, Y. Zhang, J. Chen, S. Zhang, J. Chen, X. Zhang, X. Song, Y. Pan, and Y. Ye. 2020. Deep Learning for Remote Sensing Image Analysis: A Comprehensive Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020), 2774–2797. <https://doi.org/10.1109/JSTARS.2020.2999735>