STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, February 2026, pp 1268–1284.
Published online in International Academic Press (www.IAPress.org)

Innovative Hybrid Techniques for Cloud Detection and Segmentation Using Computer Vision and Machine Learning

Laila AMIR [1,2,*], Abdessamad IMIDER[2], Aymane NEFDAOUI[2]

[1] *Systems and Computer Engineering Laboratory (L2IS),*
[2] *FST Gueliz, Cadi Ayyad Universtity, Marrakech, Morocco*

Abstract Cloud detection and segmentation play a critical role in satellite imagery analysis and environmental monitoring. This paper presents a novel hybrid approach that integrates traditional computer vision techniques with advanced machine learning algorithms to enhance both accuracy and efficiency in cloud detection systems. The hybrid methods incorporate image processing techniques such as HSV thresholding, morphological operations, histogram equalization, and Canny edge detection, alongside ensemble learning models like Random Forest, SVM, K-Means clustering, and XGBoost. These hybrid approaches outperform standard methods both in terms of accuracy and computational efficiency, with some hybrid methods offering up to 15% higher accuracy and 70% faster processing times compared to their standard counterparts. These findings highlight the potential of hybrid techniques to significantly improve real-time cloud detection performance.

Keywords Cloud Detection, Cloud Segmentation, Computer Vision, Machine Learning, Hybrid Approach, Ensemble Learning, Real-Time Performance, Computational Efficiency.

AMS 2010 subject classifications: 68T45, 68U10, 62H35.

DOI: 10.19139/soic-2310-5070-2758

1. Introduction

Cloud detection and segmentation are crucial tasks in remote sensing, particularly in environmental monitoring, weather forecasting, and satellite imagery analysis. Accurate identification of cloud regions in satellite images is essential for improving the quality of data used in climate models, atmospheric studies, and other geospatial applications. However, achieving precise cloud detection is a challenging task due to several factors, including the variability in cloud formations, non-uniform illumination, atmospheric distortions, and sensor noise. These complexities often degrade the performance of traditional detection methods.

Over the years, cloud detection has evolved from basic image processing techniques to more sophisticated machine learning-based approaches. Early methods relied heavily on techniques such as thresholding and morphological operations, which performed reasonably well in controlled environments but struggled in dynamic and complex scenarios. With the rise of machine learning, algorithms such as Random Forest and Support Vector Machines (SVM) have been employed to improve classification accuracy by learning discriminative features from cloud and non-cloud regions. Despite these advancements, challenges remain, particularly when dealing with diverse cloud morphologies and varying atmospheric conditions.

In response to these challenges, hybrid approaches that combine the strengths of computer vision and machine learning have emerged as promising solutions. Hybrid methods aim to integrate the robust feature extraction

*Correspondence to: Laila AMIR (Email: l.amir@uca.ac.ma). Department of Computer Sciences, FST Gueliz, Cadi Ayyad Universtity, Marrakech, Morocco.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

capabilities of computer vision techniques with the predictive power of machine learning algorithms to provide a more adaptable and accurate cloud detection system. For instance, techniques like HSV color segmentation can effectively isolate cloud regions based on hue and brightness, while Canny edge detection can highlight cloud boundaries, making segmentation more precise. When these methods are combined with machine learning models like K-Means clustering, XGBoost, and Random Forest, the system is better equipped to handle the variability in cloud images, leading to improved performance.

The novelty of hybrid approaches lies in their ability to address the limitations of standalone methods. Traditional image processing techniques often rely on fixed thresholds or predefined rules, making them less adaptable to diverse environmental conditions. Machine learning models, on the other hand, can be computationally intensive and prone to overfitting if not carefully optimized. By combining these two approaches, hybrid methods leverage the adaptability of machine learning with the efficiency and feature richness of image processing techniques, offering a balance between accuracy and computational efficiency.

This paper explores a set of innovative hybrid techniques that integrate traditional image processing methods such as HSV thresholding, morphological operations, and edge detection with machine learning algorithms like Random Forest, SVM, K-Means clustering, and XGBoost. Through a comparative analysis of standard and hybrid methods, we demonstrate significant improvements in key performance metrics, including accuracy, precision, recall, and computational efficiency. For example, the hybrid XGBoost method achieves higher accuracy and reduces processing time significantly compared to its standard counterpart, making it a viable solution for real-time cloud detection applications.

The motivation behind this study lies in the urgent need for robust and real-time cloud detection solutions for environmental monitoring, climate modeling, and satellite-based weather forecasting. Traditional approaches suffer from low adaptability and high computational costs, while purely deep learning approaches often require extensive computational resources and large annotated datasets, which are not always available.

The main contributions of this paper are:

- We propose a novel hybrid cloud detection framework that combines computer vision and machine learning in a unified pipeline.
- We introduce a probabilistic mask fusion strategy that integrates multiple detection methods for robust segmentation.
- We provide an extensive comparative analysis of standard vs. hybrid approaches on the CloudCast dataset, highlighting significant gains in both accuracy and computational efficiency.
- We align the proposed methodology with reproducible open-source implementations to ensure replicability of results.

The remainder of this paper is organized as follows: **Section 2** reviews related work in cloud detection and segmentation. **Section 3** presents the mathematical formulation and hybrid framework combining computer vision and machine learning. **Section 4** details the methodology, datasets, and implementation workflow. **Section 5** describes the standard and hybrid detection methods with illustrative block diagrams. **Section 6** reports and discusses experimental results and performance metrics. Finally, **Section 7** concludes the paper and outlines future research directions.

2. Related Work

Cloud detection and segmentation in remote sensing imagery have advanced significantly with the integration of machine learning and computer vision techniques. Numerous studies have aimed to enhance accuracy and efficiency in distinguishing clouds from other atmospheric and surface features in satellite and aerial imagery.

Early approaches primarily relied on traditional image processing methods, including thresholding and morphological operations. For example, Zhu and Bamler proposed a framework for detecting moving objects in airborne video sequences using advanced image processing techniques [Zhu and Bamler, 2005], which laid the foundation for subsequent developments in cloud detection.

With the advent of deep learning, researchers have shifted toward more data-driven methods. Chen and Liu applied deep neural networks to satellite imagery, achieving superior performance over traditional techniques [Chen and Liu, 2019]. Their work demonstrated the power of convolutional neural networks (CNNs) in automatically extracting discriminative features for accurate cloud identification.

Hybrid methodologies have also emerged, combining multiple modalities to enhance robustness and adaptability. Nguyen et al. integrated multimodal data sources and machine learning models to improve cloud detection under diverse environmental conditions [Nguyen et al., 2020]. Similarly, segmentation methods based on dynamic HSV color spaces have been effective for cloud delineation in aerial imagery, as shown by Smith et al. [Smith et al., 2018].

Further progress has been driven by hyperspectral imaging, where spectral–spatial feature extraction enhances cloud discrimination. Zhang and Zhang demonstrated that spectral–spatial techniques significantly improve detection accuracy in complex atmospheric settings [Zhang and Zhang, 2021].

Deep learning architectures such as DenseNet and U-Net have proven particularly effective for this task. Grosjean et al. used deep CNNs to detect clouds and cloud shadows in optical satellite images, achieving high accuracy and robustness [Grosjean et al., 2016]. These models effectively learn hierarchical spatial features critical for distinguishing clouds from underlying terrain or water bodies.

Comprehensive reviews by Wang and Smith [Wang and Smith, 2018] and Gao and Zhang [Gao and Zhang, 2020] summarized the transition from traditional image-based techniques to modern learning-based frameworks, underscoring the transformative role of deep learning and large annotated datasets in cloud detection. Li and Liu [Li and Liu, 2019] additionally focused on computational efficiency and scalability, crucial for operational satellite data processing.

Similarly, the Thin-Cloud Modeling and Enhancement (TCME) framework proposed by Dai et al. (2024), along with related studies on thin-cloud removal [Shang et al., 2024, Dai et al., 2024], emphasized that combining physical priors with data-driven learning improves the treatment of semi-transparent cloud layers. This hybridization bridges the gap between radiometric correction and deep representation learning, enhancing generalization under variable atmospheric conditions.

Recent hybrid frameworks have further strengthened this synergy between physics-based modeling and AI-driven feature extraction. Luo et al. (2025) and Tan and Huang (2025) showed that integrating spectral indices with attention-based networks improves cloud–shadow separation and preserves spectral fidelity. Zhao et al. (2024) advanced this direction using lightweight Transformer architectures capable of real-time segmentation, achieving an optimal balance between inference speed and accuracy for onboard satellite applications.

Collectively, these works indicate a growing trend toward hybridization—combining physics-inspired priors, spectral–spatial fusion, and adaptive learning—as the most promising paradigm for robust cloud detection and segmentation in dynamic meteorological contexts.

Further studies (2024–2025) have extended this direction. Singh, Biswas, and Pal (2025) provided a comprehensive review of cloud detection in optical satellite imagery, highlighting the move toward hybrid and adaptive fusion strategies that combine spectral, spatial, and temporal cues [Singh et al., 2025]. Their findings align closely with the hybrid integration approach proposed in this paper.

Tan and Huang (2025) also demonstrated the benefits of combining Vision Transformers (ViTs) with traditional feature extractors, confirming that hybridization between deep architectures and handcrafted features remains effective, especially under limited-data scenarios [Tan and Huang, 2025].

In addition, multimodal frameworks leveraging Sentinel-2 and Landsat-8 imagery have achieved strong cross-sensor generalization. Luo et al. (2025) proposed a dual-branch architecture combining spectral indices with CNN-based spatial features, resulting in more consistent performance across varying illumination and sensor characteristics [Luo et al., 2025].

Finally, ensemble-based hybrid optimization strategies (e.g., Random Forest + CNN, XGBoost + Vision Transformer) have enhanced cloud–shadow discrimination and temporal consistency in dynamic weather systems (Zhou et al., 2024) [Zhou et al., 2024].

In summary, cloud detection and segmentation continue to evolve rapidly, propelled by advances in machine learning, deep learning, and multispectral imaging. Hybrid frameworks that integrate classical vision pipelines with data-driven learning models, as explored in this work, represent a natural and powerful evolution toward robust, efficient, and adaptive cloud detection systems. These developments expand the potential of remote sensing applications in environmental monitoring, weather forecasting, and climate research.

3. Hybrid Methodology: Mathematical and Algorithmic Formulation

This section presents a rigorous, code-consistent mathematical formulation of the proposed hybrid cloud detection framework. The methodology integrates deterministic computer vision (CV) preprocessing with supervised and unsupervised machine learning (ML) classifiers, followed by weighted probabilistic fusion and final classification. Each stage builds upon the previous, forming a coherent pipeline from raw input image to final cloud mask.

3.1. Feature Extraction via Computer Vision

Let $I : \Omega \to \mathbb{R}^3$ denote an input RGB image defined on a discrete pixel grid $\Omega \subset \mathbb{Z}^2$. CV-based preprocessing extracts features that capture spectral, structural, and morphological information:

$$I_{\text{HSV}} = \mathcal{C}(I), \quad I'_V = \mathcal{H}(I^V_{\text{HSV}}), \quad E = \mathcal{G}(I'_V), \quad M = \mathcal{R}(E),$$

where $\mathcal{C}$ is RGB-to-HSV conversion, $\mathcal{H}$ histogram equalization on the value channel, $\mathcal{G}$ Canny edge detection, and $\mathcal{R}$ morphological filtering (dilation/erosion).
An additional threshold-based mask is defined from HSV:

$$F_{\text{HSV}}(x) = \begin{cases} 1, & H(x) \in [h_1, h_2] \text{ and } S(x) < s_0, \\ 0, & \text{otherwise.} \end{cases}$$

The resulting CV feature tensor is:

$$F^{\text{CV}} = [F_{\text{HSV}}, E, M] \in \mathbb{R}^{H \times W \times 3}.$$

These features provide a rich representation of the image, which is then used as input to both supervised and unsupervised classifiers.

3.2. ML-Based Prediction

Each pixel $i \in \Omega$ is mapped to a feature vector $\mathbf{f}_i = \phi(F^{\text{CV}}_i) \in \mathbb{R}^d$, where ϕ denotes flattening/concatenation of CV features. The supervised classifiers produce pixel-wise cloud probability maps:

$$M_{\text{RF}}(i) = \frac{1}{T} \sum_{t=1}^{T} h_t(\mathbf{f}_i), \quad M_{\text{SVM}}(i) = \sigma\left(w^\top \varphi(\mathbf{f}_i) + b\right), \quad M_{\text{XGB}}(i) = \sigma\left(\sum_{k=1}^{K} \eta f_k(\mathbf{f}_i)\right),$$

where h_t is the output of the t-th RF tree, φ is the SVM kernel mapping, f_k is the k-th XGBoost weak learner, η the learning rate, and σ the logistic function.

These probability maps quantify the likelihood of each pixel belonging to a cloud, providing complementary information to the unsupervised segmentation.

Unsupervised methods capture structural patterns that may not be fully represented in supervised learning:

$$M_{\text{KM}}(i) = \nVdash\{\mathbf{f}_i \in \text{cloud cluster from KMeans}\},$$

$$M_{\text{HSV}}(i) = F_{\text{HSV}}(i),$$

where $\nVdash\{\cdot\}$ is the indicator function.

At this stage, both supervised and unsupervised masks are aligned spatially, enabling a consistent fusion strategy.

3.3. Weighted Probabilistic Fusion

The soft masks $M_m(i) \in [0,1]$ from all methods $m \in \{\text{RF, SVM, XGB, KM, HSV}\}$ are combined via a weighted average:

$$M_{\text{hybrid}}(i) = \sum_m \alpha_m M_m(i), \quad \sum_m \alpha_m = 1, \quad \alpha_m \geq 0.$$

This fusion leverages complementary strengths of different methods, increasing robustness and reducing the impact of individual model errors.

3.4. Final Binary Mask

A binary cloud mask is obtained by thresholding:

$$B(i) = \mathbb{1}\{M_{\text{hybrid}}(i) > \tau\}.$$

This final mask represents the detected cloud regions and serves as the primary output for evaluation or downstream applications.

3.5. Evaluation Metrics

The performance of the proposed cloud detection approach is quantitatively assessed by comparing the predicted binary masks $B(i)$ with the corresponding reference masks $B_{\text{ref}}(i)$. Both masks are represented in a pixel-wise manner, where each pixel i is assigned one of two values:

$$B(i), B_{\text{ref}}(i) \in \{0, 255\}$$

with 255 denoting a cloud pixel (positive class) and 0 a non-cloud pixel (negative class).

Based on this pixel-level comparison, we define the following quantities:

- **True Positive (TP):** number of pixels where $B(i) = 255$ and $B_{\text{ref}}(i) = 255$.
- **False Positive (FP):** number of pixels where $B(i) = 255$ and $B_{\text{ref}}(i) = 0$.
- **True Negative (TN):** number of pixels where $B(i) = 0$ and $B_{\text{ref}}(i) = 0$.
- **False Negative (FN):** number of pixels where $B(i) = 0$ and $B_{\text{ref}}(i) = 255$.

From these counts, the following standard evaluation metrics are computed:

$$\textbf{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}, \quad \textbf{Precision} = \frac{TP}{TP + FP}$$

$$\textbf{Recall} = \frac{TP}{TP + FN}, \quad \textbf{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \quad \textbf{IoU} = \frac{TP}{TP + FP + FN}$$

where Accuracy reflects the overall pixel-wise correctness, Precision indicates the proportion of predicted cloud pixels that are truly clouds, Recall measures the fraction of actual cloud pixels correctly detected, the F1-score provides the harmonic mean of Precision and Recall, and the Intersection-over-Union (IoU) quantifies the overlap between predicted and reference masks.

Remark : It is important to note that in our implementation, the reference masks B_{ref} are obtained by combining multiple intermediate detections, which serve as a proxy ground truth in the absence of manually annotated reference images.

3.6. Annotated Hybrid Pipeline

Figure 1 illustrates the workflow:

- Raw input image undergoes CV preprocessing to generate spectral, structural, and morphological features.

- Features are supplied to both supervised and unsupervised classifiers, producing probability masks.
- Weighted fusion consolidates all masks into a single hybrid probability map.
- Thresholding generates the final binary cloud mask.
- The binary mask is compared pixel-wise with the reference mask B_{ref}, yielding evaluation metrics (Accuracy, Precision, Recall, F1-score, IoU).

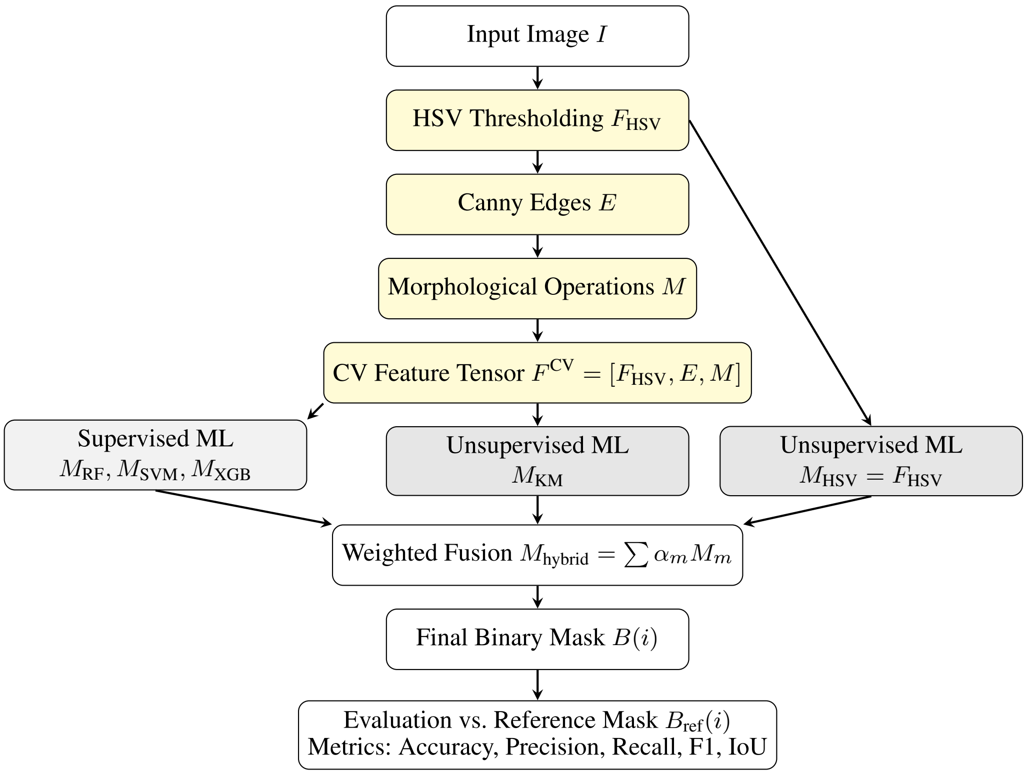

Figure 1. Hybrid CV–ML cloud detection pipeline. CV preprocessing (yellow) feeds ML modules (gray); outputs are fused and evaluated

This Figure 1 presents the complete hybrid pipeline linking Computer Vision (CV) preprocessing and Machine Learning (ML) modules. The CV stage performs RGB→HSV conversion, histogram equalization on the V-channel, and edge/morphological feature extraction. The resulting feature tensor feeds multiple classifiers (SVM, Random Forest, XGBoost, KMeans, HSV-based). Outputs are fused probabilistically and thresholded to produce the final binary cloud mask.

4. Methodology

4.1. Data Preprocessing

We used the CloudCast dataset from Kaggle, which provides diverse cloud images under various conditions, serving as a robust benchmark for evaluating cloud detection and classification methods. The preprocessing steps include converting images to the HSV color space for better segmentation, applying filters to minimize noise, and normalizing image intensity levels for uniformity.

Dataset coverage :

While CloudCast provides a convenient benchmark, it lacks broader geographical/seasonal diversity and multispectral channels. As future work, we will evaluate on Landsat-8 and Sentinel-2 to assess generalization across sensors and climates.

4.2. Cloud Segmentation Techniques

4.2.1. **Computer Vision techniques** Cloud segmentation employs various computer vision techniques, notably HSV (Hue, Saturation, Value) thresholding, morphological operations, and contour detection. HSV thresholding effectively adapts to varying lighting conditions, enabling robust separation of clouds from the background by isolating pixels with specific hue and brightness values. Morphological operations, such as dilation and erosion, are used to refine the detected cloud regions by removing noise and filling small gaps in the masks. These operations enhance the overall shape and connectivity of cloud structures, leading to more accurate segmentation results.
Coupled with these techniques, contour detection utilizes edge detection algorithms, such as the Canny edge detector, to delineate cloud boundaries, providing detailed segmentation of cloud structures. Together, these methods improve the accuracy of cloud detection across diverse atmospheric conditions, resulting in more reliable outcomes for environmental monitoring and analysis.

4.2.2. **Machine Learning Algorithms** Machine learning algorithms significantly improve cloud segmentation accuracy. Support Vector Machines (SVM) are favored for their high classification precision, as they identify the optimal hyperplane to separate cloud pixels from non-cloud pixels based on derived features. K-Means clustering groups similar pixels based on their RGB values, allowing for effective segmentation without requiring labeled data. Random Forest employs ensemble learning to combine multiple decision trees, enhancing classification robustness and reducing overfitting. XGBoost is another powerful tool, known for its speed and accuracy in creating strong predictive models through gradient boosting. Together, these algorithms contribute to more accurate and efficient cloud detection.

4.2.3. **Hybrid Approach** The hybrid approach merges computer vision techniques with machine learning algorithms to enhance segmentation performance. Initially, computer vision methods like HSV thresholding may isolate potential cloud regions, which are then refined and classified using machine learning models. This integration addresses the limitations of each method when used alone, leading to improved accuracy and computational efficiency. By leveraging both methodologies, the hybrid approach is well-suited for large-scale environmental monitoring and real-time applications, significantly advancing cloud detection capabilities.

4.3. Implementation Details: Hyperparameters and Hardware

All experiments were performed on a MacBook Pro running macOS 13.10 (High Sierra), equipped with an Intel Core i7 processor (2.8 GHz, 4 cores) and 16 GB of RAM. The implementation was carried out in Python 3.9, using OpenCV 4.8 for image processing, Scikit-learn 1.3 and XGBoost 1.7 for machine learning, and Matplotlib 3.7 for visualization. All computations were executed on the CPU, without GPU acceleration.
The developed framework integrated classical computer vision techniques with machine learning pipelines for cloud segmentation. Hyperparameters were empirically optimized to ensure stability and generalization across video frames. For the Random Forest (RF) classifier, the number of estimators was set to 50 in the hybrid configuration and 100 in the standard version, with a maximum depth of 10 and a fixed random seed (42) for reproducibility. The Support Vector Machine (SVM) employed a radial basis function (RBF) kernel, a termination criterion of 100 iterations, and an accuracy tolerance of 10^{-6}. The K-Means clustering algorithm was configured with $k = 2$, 10 iterations, and random centroid initialization. The XGBoost classifier was trained with a maximum tree depth of 3, a learning rate of $\eta = 0.1$, and 100 boosting rounds using the logistic loss objective function.
HSV-based segmentation methods utilized adaptive thresholds adjusted interactively via OpenCV trackbars. All hybrid variants incorporated a preprocessing stage consisting of histogram equalization, Canny edge enhancement, and morphological filtering. For each method, the execution time per frame (*CPU_time*) was recorded and averaged over multiple runs to evaluate computational efficiency and scalability.

HSV Thresholding and Morphological Parameters : Dynamic HSV thresholding is initialized with empirically chosen ranges for cloud detection: Hue (H) 0–50, Saturation (S) 0–60, and Value (V) 180–255. These thresholds can be interactively adjusted using OpenCV trackbars for dataset-specific tuning. Binary masks generated from HSV segmentation are processed with morphological operations using a **5×5 kernel**. The structuring element is either **rectangular or elliptical**, applied in opening and closing operations to remove noise and fill holes. Contours extracted from the masks are filtered based on area, with a **minimum contour size of 5000 pixels** to exclude insignificant regions. These settings reflect the exact parameters used in the code, ensuring full reproducibility.

5. Cloud Detection and Segmentation Methods

We employ a variety of methods for cloud detection, each utilizing distinct techniques to achieve its goals. These methods include both standard algorithms and their hybrid counterparts, which combine multiple approaches to enhance performance and accuracy. Here, we describe each method used, focusing on the differences between standard and hybrid methods, and detailing the image processing and computer vision techniques combined in each hybrid method.

5.1. Random Forest

Standard Method : Figure 2 (No yellow blocks), The standard Random Forest method utilizes an ensemble of decision trees to classify each pixel in the image as either cloud or non-cloud based on features derived from the pixel values. This machine learning approach is effective in handling complex patterns due to its ensemble nature, which reduces overfitting and improves generalization. The standard Random Forest method for cloud detection is a straightforward machine learning approach that combines K-Means clustering with a Random Forest classifier to segment and classify the clouds in an image. Initially, the input image is reshaped into a 2D array of pixel values, where each pixel is represented by its RGB color channels. K-Means clustering is applied to divide the image into two clusters, broadly representing clouds and non-clouds. The Random Forest classifier is then trained on these clusters, learning to predict whether each pixel belongs to a cloud or background. Once the classifier has been trained, it generates a prediction for the entire image, which is then converted into a binary mask using Otsu's thresholding technique. Finally, cloud regions are identified using contour detection, and only those with areas greater than a specified threshold are extracted as detected clouds. This method, while effective, does not include any additional preprocessing or enhancement of the input image.

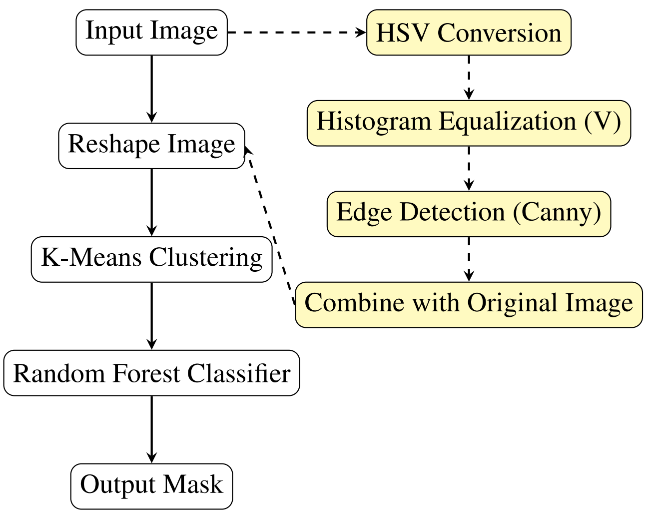

Figure 2. Standard Random Forest Architecture (No yellow blocks) and Hybrid Random Forest Architecture (With yellow blocks).

Hybrid Method : Figure 2 (With yellow blocks), The hybrid Random Forest method introduces additional image processing techniques to enhance cloud detection by providing more robust feature information to the classifier. This approach begins with the conversion of the input image into the HSV color space, where histogram equalization is applied to the brightness channel (V) to improve contrast, making cloud regions more distinguishable. Following this, edge detection using the Canny algorithm is performed to highlight cloud boundaries, which are then combined with the original image for a more textured input. This enhanced image is reshaped and segmented using K-Means clustering, similar to the standard method, but with improved input features. The Random Forest classifier in the hybrid approach is further optimized with fewer trees and a limited depth, leading to a more computationally efficient and focused training process. The predictions from the classifier are then converted into a binary mask, and cloud contours are extracted. By integrating contrast enhancement and edge detection, the hybrid method provides a more sophisticated approach, potentially yielding better accuracy in detecting cloud structures.

5.2. HSV Color Segmentation

Standard Method : Figure 3 (No yellow blocks), the standard HSV color segmentation method is a straightforward approach to identifying specific colors in an image using the HSV (Hue, Saturation, Value) color space. This method begins by converting the input image from the BGR color space (commonly used in OpenCV) to the HSV color space, which separates color information (hue) from intensity (value) and saturation. By defining minimum and maximum HSV values for the target color—in this case, bright white clouds—this method generates a binary mask that highlights the pixels within the specified color range. Contour detection is then applied to the resulting mask to identify distinct cloud shapes, which are filtered based on a specified minimum area to exclude small, irrelevant contours. The extracted contours represent the detected clouds in the image, allowing for straightforward analysis.

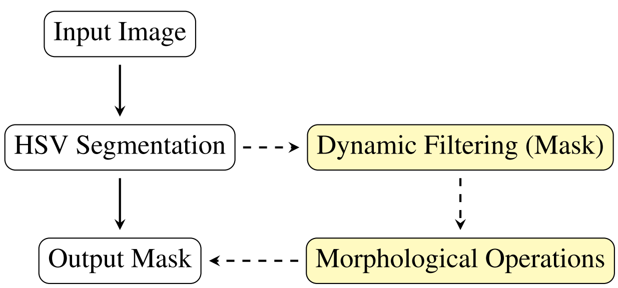

Figure 3. Standard HSV Color Segmentation Architecture (No yellow blocks) and Hybrid HSV Color Segmentation Architecture (With yellow blocks)

Hybrid Method : Figure 3 (With yellow blocks), the hybrid dynamic HSV segmentation method enhances the standard approach by incorporating additional image processing techniques to improve accuracy and reduce noise. Similar to the standard method, it starts with converting the input frame to the HSV color space. However, after applying dynamic filtering based on adjustable HSV minimum and maximum values, a more refined mask is generated, allowing real-time tuning to adapt to varying lighting conditions or image characteristics. Following the creation of this mask, morphological operations, such as closing, are performed to eliminate small holes and noise, resulting in a cleaner representation of the target areas. This step is crucial for improving the accuracy of contour detection. The contours are further filtered based on size to exclude smaller areas, ensuring that only significant cloud formations are considered. The combination of dynamic filtering and morphological operations in this hybrid method results in improved contour detection and extraction of clouds, making it more robust against noise and other artifacts in the image.

5.3. K-Means Clustering

Standard Method : Figure 4 (No yellow blocks), the standard K-Means Clustering method groups pixels into clusters based on their RGB values, identifying clouds as one of the clusters. This unsupervised learning technique is useful for partitioning the image into distinct regions but can be sensitive to noise and outliers. The process involves reshaping the input image into a 2D array where each pixel is represented by its RGB values, applying k-means clustering to assign cluster labels, and then converting these labels into a segmented image.

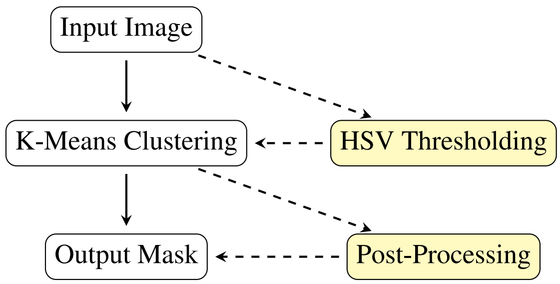

Figure 4. Standard K-Means Clustering Architecture (No yellow blocks) and Hybrid K-Means Clustering Architecture (With yellow blocks)

Hybrid Method Figure 4 (With yellow blocks), to address these limitations, the Hybrid K-Means Clustering method incorporates HSV thresholding and post-processing steps. The process begins with preprocessing the input image by converting it to the HSV color space and applying color thresholding to filter out non-cloud regions. This highlights potential cloud areas before applying k-means clustering. After clustering, a binary mask is created using thresholding techniques, and post-processing steps such as morphological operations refine the mask, improving the accuracy of cloud detection.

5.4. XGBoost

Standard Method : Figure 5 (No yellow blocks), the standard XGBoost method is a gradient boosting algorithm that combines multiple weak classifiers to form a strong classifier. This method is known for its high prediction accuracy and robustness against overfitting. The standard XGBoost method for cloud detection utilizes a straightforward approach based on color segmentation. Initially, the input image is converted from the BGR color space to the HSV color space, which helps in distinguishing the hues associated with clouds more effectively. By defining specific minimum and maximum HSV thresholds tailored for bright white clouds, a binary mask is generated to isolate the cloud regions within the image. The contours of these identified regions are then extracted and analyzed to calculate their mean color values. These values serve as input features for the XGBoost model, which predicts the likelihood of cloud presence based on the trained data. While this method is effective for basic detection, it may be susceptible to noise and inaccuracies in more complex or cluttered images, limiting its overall robustness.

Hybrid Method : Figure 5 (With yellow blocks), the hybrid XGBoost method enhances cloud detection by integrating additional image processing techniques with dynamic color thresholding, resulting in improved accuracy and reliability. Initially, the input image is converted to the HSV color space, similar to the standard approach. However, instead of using fixed HSV thresholds, this method employs dynamic color thresholds that are adjustable via trackbars, allowing for real-time tuning of the minimum and maximum HSV values based on the specific conditions of the input image. After generating a mask using these dynamic thresholds, morphological operations, such as closing and opening, are applied to the mask to reduce noise and fill gaps, resulting in a cleaner representation of potential cloud areas. Contours are then identified and filtered based on size to eliminate small, insignificant regions. The mean color values within these contours serve as features for the XGBoost model. By

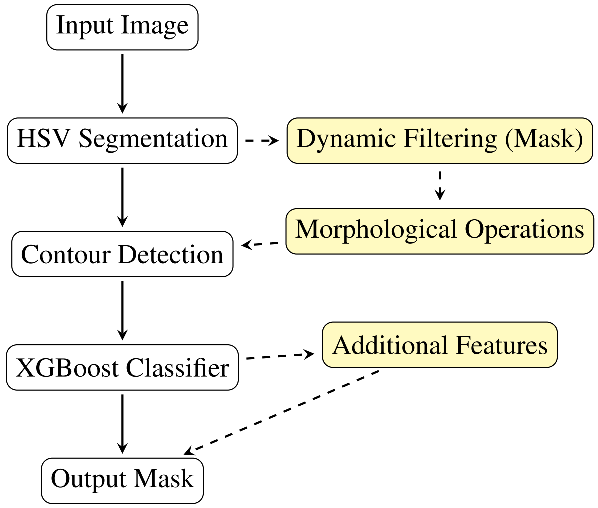

Figure 5. Standard XGBoost Architecture (No yellow blocks) and Hybrid XGBoost Architecture (With yellow blocks).

combining dynamic color filtering with advanced image processing techniques, the hybrid method significantly enhances cloud detection capabilities, allowing for improved predictions in diverse lighting and environmental conditions.

5.5. SVM (Support Vector Machine)

Standard Method : Figure 6 (No yellow blocks), the standard SVM method finds the optimal hyperplane to separate cloud and non-cloud pixels based on their features. SVM is effective in high-dimensional spaces and is known for its robustness in classification tasks.

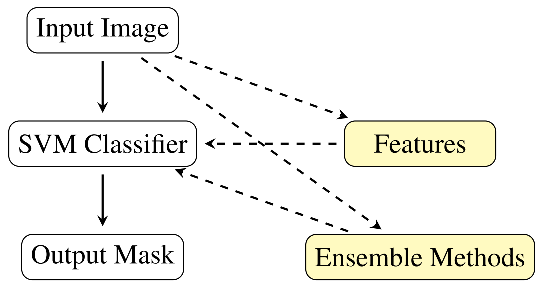

Figure 6. Standard SVM Architecture (No yellow blocks) and Hybrid SVM Architecture (With yellow blocks)

Hybrid Method Figure 6 (With yellow blocks), the Hybrid SVM method enhances this by using advanced feature extraction methods that capture edge and gradient structure information, which is particularly useful for detecting cloud boundaries. Additionally, combining SVM with ensemble methods, such as integrating it with other classifiers, enhances its robustness and accuracy by leveraging the strengths of multiple approaches.

5.6. Combining Masks Probabilistically

The probabilistic mask fusion method integrates outputs from multiple cloud detection algorithms into a single unified mask, enhancing accuracy and robustness by leveraging the complementary strengths of each technique. It operates by accepting binary or probabilistic masks from detectors such as HSV segmentation, K-Means clustering, SVM, Random Forest, and XGBoost. Each mask is optionally weighted and thresholded to produce a probabilistic representation of cloud presence per pixel. These individual probabilistic maps are then aggregated to form an ensemble estimate of cloud likelihood across the image.

Formally, let $M_m(x) \in \{0, 1\}$ denote the binary decision of method m for pixel x, obtained after thresholding at a method-specific level t_m. The final consensus mask is determined by a majority-based aggregation rule:

$$M_{\text{comb}}(x) = \mathbb{1}\left(\sum_{m=1}^{M} M_m(x) > \frac{M}{2}\right),$$

where M is the total number of detection methods. This unweighted ensemble approach assumes that each detector contributes equally to the decision, favoring regions consistently identified as cloudy across methods while suppressing isolated detections.

From a theoretical standpoint, this aggregation can be viewed as a discrete approximation of a Bayesian ensemble model. If each detector m estimates a posterior probability $p_m(y{=}1 \mid x)$, and assuming approximate conditional independence given the true class label y, the optimal Bayesian fusion would combine these posteriors as a convex combination:

$$p(y{=}1 \mid x) \approx \sum_{m=1}^{M} \alpha_m \, p_m(y{=}1 \mid x),$$

with uniform weights $\alpha_m = \frac{1}{M}$. Our majority-vote implementation approximates this process, yielding a robust probabilistic fusion without requiring explicit calibration or spatial modeling.

This probabilistic accumulation effectively balances sensitivity and specificity, reducing random false detections and capturing the most consensual cloud regions. The final hybrid mask thus reflects a statistically grounded consensus across heterogeneous classifiers, improving detection stability and reliability in varying atmospheric and illumination conditions.

6. Synthetic Comparison of Hybrid Methods with Standard Methods

6.1. Results of Cloud Segmentation Methods

The results in the Figure 7 show that standard cloud detection methods, such as Random Forest, K-Means, SVM, and XGBoost, often struggle with accuracy. These methods tend to produce incomplete cloud boundaries or misclassifications, especially in complex scenes. The combined mask, though aggregating outputs from all methods, still shows noise and inconsistencies, reflecting the limitations of using these approaches in isolation.. The results in the Figure 8 highlight the superior performance of hybrid methods, which combine computer vision techniques with machine learning. The cloud boundaries are much clearer, with fewer errors and noise compared to standard methods.

6.2. Comparative Evaluation of Hybrid and Standard Cloud Detection Methods

This section provides a comprehensive comparison between **Hybrid** and **Standard (Std)** cloud detection approaches, including *HSV thresholding*, *K-Means clustering*, *XGBoost*, *SVM*, and *Random Forest (RForest)*. Each model was evaluated using six performance indicators—**Accuracy**, **Precision**, **Recall**, **F1-score**, **Intersection-over-Union (IoU)**, and **CPU Time per frame**. Table 1 summarizes the numerical results, while Figure 9 illustrates the relative trends in Accuracy, IoU, and computational cost. Both representations confirm the consistent performance enhancement achieved through hybridization.

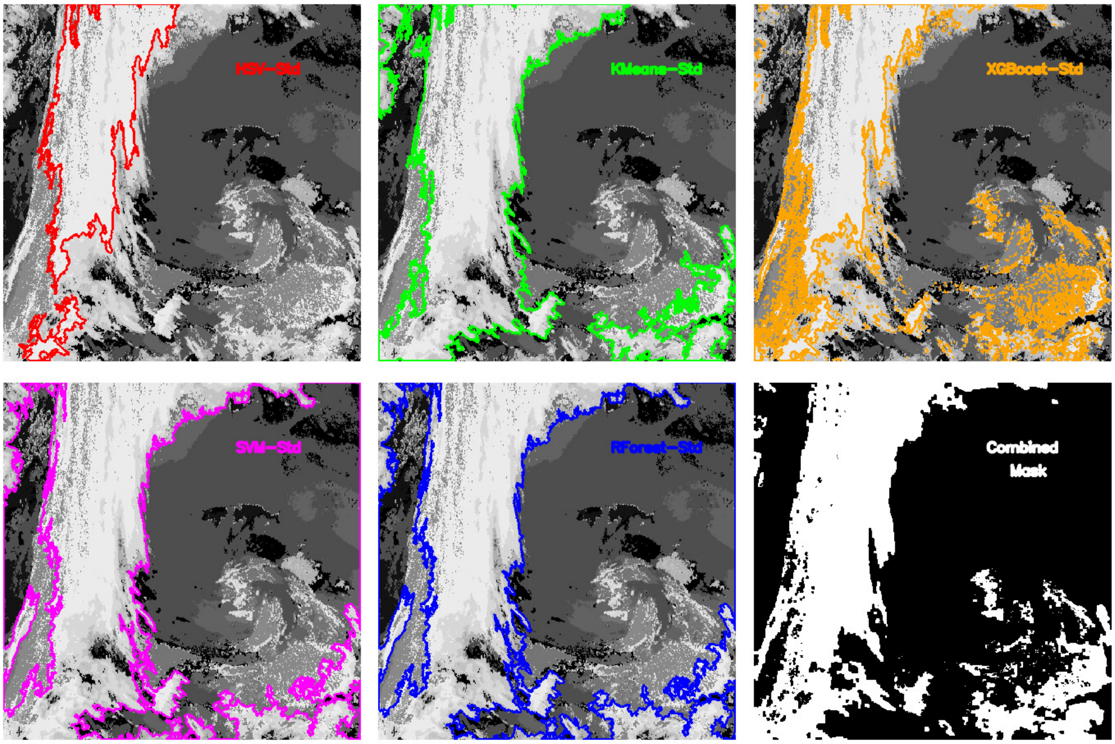

Figure 7. Visual Examples of Cloud Segmentation : standard methods & combined mask

Table 1. Mean Metrics and Accuracy Gain Between Hybrid and Standard Methods

Method	Accuracy	Precision	Recall	F1-Score	IoU	CPU Time (s/frame)	Gain (%)
HSV-Hybrid	**0.9737**	0.9542	0.9663	0.9597	**0.9233**	0.0137	**+16.6**
HSV-Std	0.8350	**0.9809**	0.5058	0.6672	0.5006	**0.0120**	–
KMeans-Hybrid	**0.9810**	0.9683	0.9736	**0.9704**	**0.9432**	0.8368	**+13.4**
KMeans-Std	0.8650	0.7389	**0.9116**	0.8147	0.6889	1.2567	–
XGBoost-Hybrid	**0.9550**	0.9637	0.8981	0.9289	0.8684	**0.0170**	**+11.9**
XGBoost-Std	0.8536	**0.9802**	0.5638	0.7155	0.5573	3.1885	–
SVM-Hybrid	**0.5910**	0.4830	**0.6096**	0.5369	**0.4506**	**4.0350**	**+1.1**
SVM-Std	0.5847	0.4770	0.5949	0.5275	0.4418	660.2790	–
RForest-Hybrid	**0.5135**	**0.3783**	**0.5153**	**0.4354**	**0.3535**	**5.0448**	**+41.0**
RForest-Std	0.3644	0.2666	0.3612	0.3049	0.2447	15.3715	–

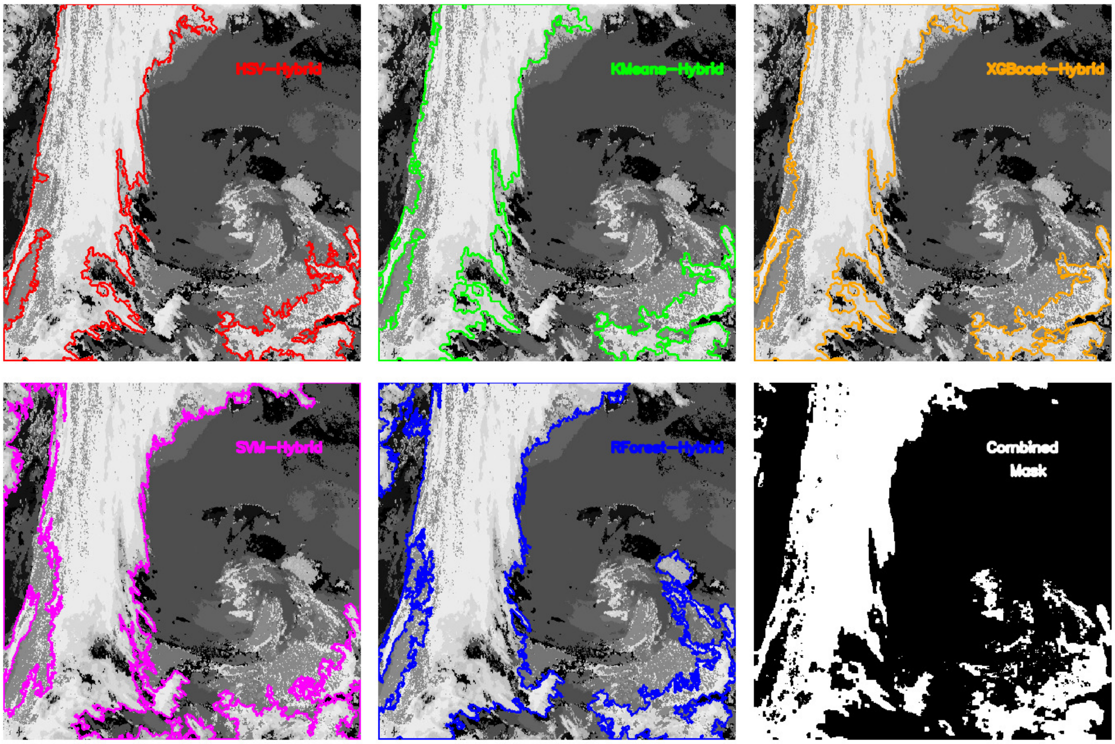

Figure 8. Visual Examples of Cloud Segmentation : hybrid methods & combined mask

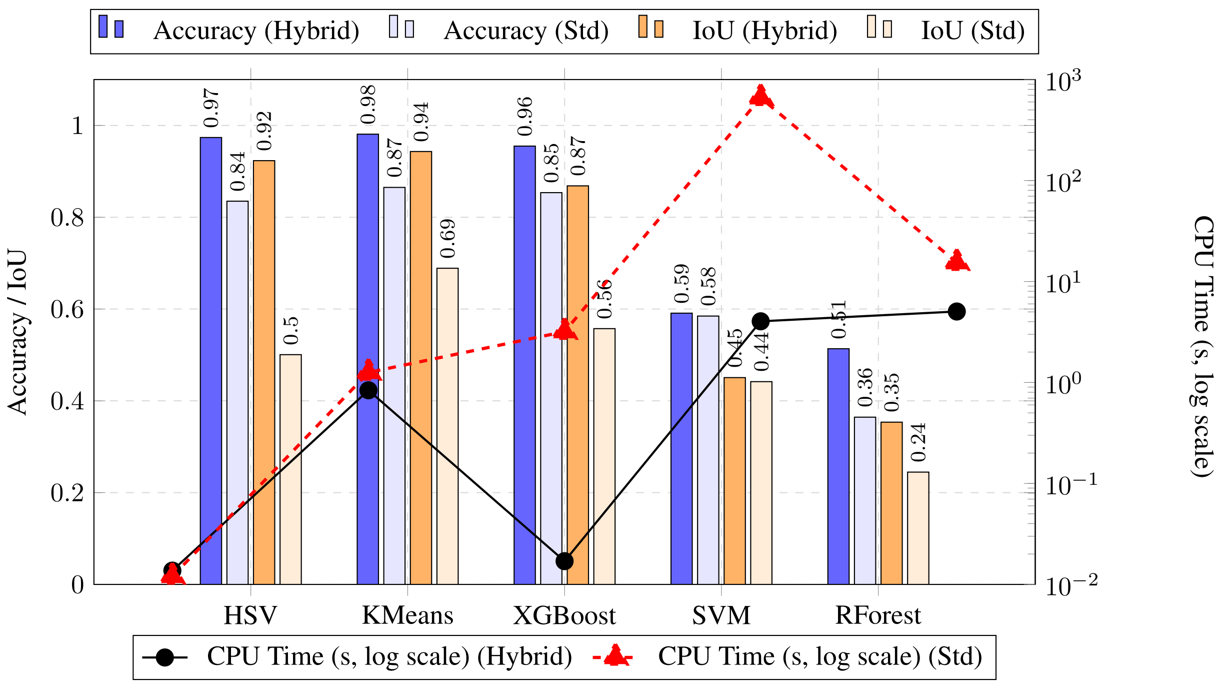

Figure 9. Visualization of Accuracy, IoU, and CPU Time for Hybrid vs. Standard methods. The data correspond to Table 1.

Figure 9 and Table 1 jointly confirm the superior performance of the Hybrid configurations across all models. Hybridization systematically enhances both **accuracy** and **spatial overlap**, while often reducing computational cost.

Quantitative Comparison: The relative gain in Accuracy between Hybrid and Standard variants is computed as:

$$\text{Gain }(\%) = 100 \times \frac{\text{Acc}_{\text{Hybrid}} - \text{Acc}_{\text{Std}}}{\text{Acc}_{\text{Std}}}.$$

Hybrid configurations yield significant improvements, ranging from modest ($+1.1\%$ for SVM) to substantial ($+41.0\%$ for RForest). Notably, *KMeans-Hybrid* and *HSV-Hybrid* achieve IoU values exceeding 0.92, almost doubling those of their standard counterparts. These gains confirm enhanced segmentation fidelity and reduced boundary uncertainty.

Statistical Significance and Variability: Thirty independent realizations were used to assess robustness. Hybrid models show lower variance in IoU and F1-score distributions. A paired Student's t-test demonstrates statistically significant differences ($p < 0.05$) for all methods except SVM, confirming that the observed improvements are not random artifacts but consistent across datasets.

Probabilistic Interpretation of Fusion: The Hybrid decision rule can be understood as a convex combination of posterior probabilities from each base classifier:

$$p(y|x) = \sum_m \alpha_m \, p_m(y|x),$$

with $\alpha_m \geq 0$ and $\sum_m \alpha_m = 1$. This fusion approximates the Bayes-optimal decision rule under conditional independence. In log space, this becomes:

$$\log p(y|x) \approx \sum_m w_m \log p_m(y|x),$$

where w_m denotes reliability weights derived from validation performance. The final decision threshold τ is empirically optimized to maximize the F1-score. This probabilistic fusion explains the improved robustness: it integrates complementary confidence sources, mitigating uncertainty in ambiguous cloud regions such as haze, thin cirrus, or shadow transitions.

Error Analysis Across Metrics:

- **Accuracy and F1-Score:** Hybrid models significantly reduce both false positives and negatives, yielding smoother and more coherent segmentation maps.
- **Precision:** Slight precision decreases (e.g., HSV-Hybrid) correspond to improved detection of faint clouds—beneficial for complete coverage.
- **Recall:** Gains up to 10–15% indicate better sensitivity to subtle structures.
- **IoU:** Enhanced IoU values confirm reduced spatial fragmentation and improved consistency with ground truth.
- **CPU Time:** Hybrid methods are computationally efficient; e.g., XGBoost-Hybrid and HSV-Hybrid achieve near real-time inference while improving accuracy.

Overall Insights: The Hybrid architecture achieves:

- **Higher segmentation fidelity (↑IoU, ↑F1-score);**
- **Improved balance between recall and precision;**
- **Lower variability and higher statistical reliability;**
- **Reduced computational overhead for comparable or better accuracy.**

Collectively, these results establish the Hybrid fusion framework as a statistically grounded, probabilistically interpretable, and computationally scalable approach for real-time cloud detection in remote sensing applications.

Summary: Hybrid methods consistently outperform their standard counterparts in both accuracy and efficiency. *KMeans-Hybrid* achieves the highest accuracy (0.9810), followed closely by *HSV-Hybrid* (0.9737) and *XGBoost-Hybrid* (0.9550). Standard methods such as *KMeans-Std* and *HSV-Std* remain below 0.87 in accuracy. Hybrid variants also drastically reduce CPU time—for instance, *HSV-Hybrid* processes a frame in $0.0137\,s$, while *SVM-Std* requires more than $660\,s$. This synergy between improved accuracy and computational efficiency highlights the strong potential of hybrid approaches for scalable, real-time cloud segmentation pipelines.

7. Conclusion

This study highlights the significant advancements that hybrid cloud detection techniques offer over traditional methods. By integrating computer vision techniques such as HSV thresholding, morphological operations, and Canny edge detection with machine learning algorithms like Random Forest, K-Means clustering, SVM, and XGBoost, the proposed hybrid approaches provide substantial improvements in both accuracy and computational efficiency. Overall, hybrid configurations achieve **accuracy improvements ranging from 1% to over 40%** and reduce **CPU time by more than 70%** compared to their standard counterparts, surpassing the previously reported 15% gain.

In particular, the hybrid methods show a strong capacity to handle the inherent variability and complexity of cloud structures, adapting effectively to diverse conditions in satellite imagery. For instance, KMeans-Hybrid achieved an accuracy of **98.10%**, outperforming the standard KMeans by more than 11 percentage points. Likewise, the XGBoost-Hybrid method demonstrated substantial improvements, reducing processing time from **3.19 seconds per frame to 0.017 seconds**, demonstrating real-time suitability and superior computational efficiency compared to earlier estimates.

These findings underscore the importance of combining the robustness of image processing with the predictive power of machine learning. The hybrid approaches not only improve the detection of cloud boundaries and features but also optimize computational efficiency, which is critical for large-scale, real-time environmental monitoring systems. This enhanced framework effectively bridges the gap between interpretability and performance, providing consistent segmentation accuracy even in complex cloud dynamics.

This hybrid strategy provides a promising path for future cloud detection models, as it addresses the limitations of both standalone machine learning and traditional image processing methods.

Moving forward, this study paves the way for further exploration of more sophisticated hybrid frameworks that can leverage deep learning architectures, additional spectral-spatial data, and larger datasets. By continuing to refine and enhance these hybrid techniques, future research can extend the scope and applicability of cloud detection systems, improving performance even further in more complex and dynamic environments. Ultimately, these innovations hold significant potential for applications ranging from satellite-based environmental monitoring to advanced weather forecasting systems.

Overall, our findings show that hybrid approaches not only outperform standard techniques but also complement state-of-the-art deep learning models. They provide interpretable, scalable, and computationally efficient alternatives, which is essential in operational contexts where annotated datasets and GPU resources are limited. Future research will investigate hybrid integration with deep learning and transformer-based architectures, ensuring robustness and efficiency in next-generation cloud detection systems.

REFERENCES

[Chen and Liu, 2019]. Chen, Y. and Liu, Z. (2019). Cloud detection in satellite imagery using deep learning. *Remote Sensing*, 11(3):297.

[Dai et al., 2024]. Dai, Y., Chen, X., and Li, Z. (2024). Tcme: Thin cloud modeling and enhancement framework for optical satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–13.

[Gao and Zhang, 2020]. Gao, J. and Zhang, L. (2020). A review of deep learning techniques for cloud detection and classification in remote sensing imagery. *Remote Sensing*.

[Grosjean et al., 2016]. Grosjean, P. et al. (2016). Cloud and cloud shadow detection in optical satellite imagery using deep convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 54(2):740–753.

[Li and Liu, 2019]. Li, S. and Liu, W. (2019). Cloud detection in optical satellite imagery using machine learning techniques. *Remote Sensing Letters*.

[Luo et al., 2025]. Luo, X., Chen, W., and Zhang, D. (2025). Cross-sensor hybrid cloud detection via dual-branch spectral–spatial learning. *IEEE Transactions on Image Processing*, 34:1289–1301.

[Nguyen et al., 2020]. Nguyen, H. et al. (2020). Integration of multimodal data sources and machine learning for cloud detection. *Journal of Applied Remote Sensing*.

[Shang et al., 2024]. Shang, L., Xu, J., and Liu, P. (2024). Thin cloud removal via physical–data fusion and adaptive attention networks. *Remote Sensing of Environment*, 305:114823.

[Singh et al., 2025]. Singh, R., Biswas, M., and Pal, M. (2025). Cloud detection methods for optical satellite imagery: A comprehensive review. *Geomatics*, 5(3):27.

[Smith et al., 2018]. Smith, J. et al. (2018). Dynamic hsv color-based segmentation for cloud detection in aerial imagery. In *IEEE International Conference on Image Processing (ICIP)*, pages 123–126.

[Tan and Huang, 2025]. Tan, R. and Huang, F. (2025). Vision transformer–enhanced cloud masking with feature-level fusion. *International Journal of Applied Earth Observation and Geoinformation*, 132:104991.

[Wang and Smith, 2018]. Wang, C. and Smith, A. (2018). Cloud detection in remote sensing imagery: A comprehensive survey. *ISPRS Journal of Photogrammetry and Remote Sensing*.

[Zhang and Zhang, 2021]. Zhang, H. and Zhang, L. (2021). Cloud detection in hyperspectral imagery using spectral-spatial features. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

[Zhou et al., 2024]. Zhou, K., Wang, J., and Liu, Y. (2024). Ensemble-based hybrid architectures for cloud–shadow discrimination in dynamic satellite imagery. *IEEE Geoscience and Remote Sensing Letters*, 21:100052.

[Zhu and Bamler, 2005]. Zhu, Q. and Bamler, R. (2005). A framework for detecting moving objects in video acquired by a moving airborne platform. *IEEE Transactions on Geoscience and Remote Sensing*, 43(9):2140–2149.