AVE Trends in Intelligent Computing Systems

Interpretable and Pattern Aware Cloud Segmentation Using Feature Semantic Learning

A. Sandhya[1], M. Kiruthigha[2], G. Deena[3,*], R. Sethuraman[4], M. Thamizharasi[5]

[1,3]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[2]Department of Information Technology, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[4]Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India.
[5]Department of Information Technology, Saveetha Engineering College, Chennai, Tamil Nadu, India.
sandhyaa@srmist.edu.in[1], kiruthir5@srmist.edu.in[2], deenag@srmist.edu.in[3], sethuraman.cse@sathyabama.ac.in[4], thamizharasi@saveetha.ac.in[5]

*Corresponding author

Abstract: The Landsat archive is essential for Earth surface monitoring and land cover changes. To utilise this wealth of data effectively, cloud and cloud shadow interference must be filtered, particularly for automated systems that process hundreds of photographs without human input. Cloud and cloud shadow interference often obscure clear ground views. Clouds block optical satellites from tracking the globe. These clouds must be precisely located for remote sensing using satellite picture archives. Despite improvements, cloud identification remains challenging, particularly in bright environments and thin cloud layers. Recent breakthroughs in deep learning-based cloud masking have improved cloud identification accuracy. These approaches are often evaluated differently, and studies compare them to threshold-based systems inconsistently. We propose Ed-CNN, a deep convolutional neural network (DCNN) with an enhanced encoder-decoder architecture, for cloud and snow segmentation. This method enhances segmentation accuracy by augmenting the encoder with the Atrous Spatial Pyramid Pooling module and the decoder by integrating information from multiple encoder stages. We recommend standardising datasets and protocols for cloud detection model benchmarking, as deep learning models trained on similar sensor data can benefit new satellites and heavily rely on training and testing datasets.

Keywords: Cloud Detection; Deep Convolutional Neural Network; Operational Land Imager; Thermal Infrared Sensor; U-Net and Semantic Learning; Satellite Images; Sensor Data.

Cite as: A. Sandhya, M. Kiruthigha, G. Deena, R. Sethuraman, and M. Thamizharasi, "Interpretable and Pattern Aware Cloud Segmentation Using Feature Semantic Learning," *AVE Trends in Intelligent Computing Systems*, vol. 2, no. 2, pp. 87–98, 2025.

Journal Homepage: https://www.avepubs.com/user/journals/details/ATICS

Received on: 16/09/2024, **Revised on:** 13/12/2024, **Accepted on:** 21/01/2025, **Published on:** 07/06/2025

DOI: https://doi.org/10.64091/ATICS.2025.000133

1. Introduction

Satellite imagery, particularly data acquired from missions such as Landsat-8, plays a crucial role in studying and monitoring the Earth's surface. The Landsat-8 satellite is outfitted with two advanced sensors, the Operational Land Imager (OLI) and the

Thermal Infrared Sensor (TIRS), both designed to capture high-resolution data across various spectral bands. This diverse spectral information is instrumental in identifying and analysing changes in terrestrial features over time [1]. The rich data provided by Landsat-8 imagery serves as a crucial resource for numerous scientific and environmental applications that include (i) Ecological Monitoring: Facilitates the observation of forests, aiding in the detection of deforestation trends, assessment of vegetation health, and tracking biodiversity patterns. (ii) Climate Change Studies: Enables the examination of temperature fluctuations, monitoring of glacial retreat, and analysis of atmospheric transformations linked to climate change. (iii) Land Cover Analysis: Supports comprehensive mapping efforts, including the expansion of urban landscapes, evaluation of forest density, and monitoring of agricultural land use [11]. Despite the powerful observational capabilities of Landsat-8, environmental factors such as cloud cover and snow cover pose significant challenges to its use. These natural phenomena can obscure the Earth's surface, impeding accurate data collection and interpretation [2]; [6]. Overcoming these obstacles is essential for ensuring the reliability and precision of satellite-derived insights. Cloud cover obscures the Earth's surface, making it difficult to accurately interpret ground-level features. This obstruction often results in data gaps or errors during interpretation.

Similarly, snow cover poses a significant challenge as it reflects sunlight in a way that resembles certain cloud formations. This similarity makes it difficult to distinguish between snow and clouds using traditional spectral data. These factors have a substantial impact on data interpretation, primarily in three key areas: (i) Reduced Accuracy: Cloud and snow interference can lead to incorrect classifications in land cover mapping or vegetation analysis, thereby reducing overall classification accuracy. (ii) Data Gaps: Persistent cloud cover in certain regions may limit the availability of clear imagery, delaying crucial research and environmental assessments [6]. This often results in gaps in the collected data. (iii) Misclassification: Clouds and snow can exhibit similar reflectance properties, making it challenging for automated detection algorithms to differentiate between them, resulting in incorrect classifications. To mitigate the effects of cloud and snow interference, several traditional techniques are commonly used. However, these methods have limitations, as (i) Spectral Thresholding: This method identifies specific spectral bands where clouds or snow exhibit distinct properties. While effective in some cases, it may struggle to differentiate thin clouds or mixed snow conditions, reducing its overall reliability. (2) Masking Algorithms: Tools such as the CFMask (C Function of Mask) algorithm are designed to identify and exclude cloud-covered pixels from imagery.

Although effective, these algorithms may still misclassify snow as clouds or vice versa. (3) Temporal Interpolation: This technique combines multiple images taken on different dates to reconstruct missing data caused by cloud cover. While this method can successfully fill data gaps, it requires frequent and consistent image acquisitions to achieve accurate results. Traditional methods for cloud and snow removal in satellite imagery face several limitations that can reduce their effectiveness and accuracy. These limitations become particularly evident in diverse geographic regions, especially in areas with mixed land cover types, complex terrain, or variable lighting conditions [3]. Methods that rely heavily on temperature thresholds may fail during transitional periods, such as spring thaw or autumn snow accumulation, where temperature fluctuations are frequent and unpredictable. Some traditional techniques require manual correction to enhance accuracy, which is both labour-intensive and subjective [8]. Additionally, complex algorithms, particularly when applied to large-scale imagery, may demand significant computing power and extended processing time, making the process resource-intensive. Older methods often rely on only a few spectral bands, missing crucial information that modern multi-sensor or multi-spectral techniques can provide. Moreover, combined cloud and snow conditions, or partial cloud coverage, often reduce the effectiveness of threshold-based models in mixed weather scenarios [11]. Traditional techniques frequently discard cloudy pixels without effectively reconstructing the lost data, limiting the accuracy and completeness of time-series analysis.

Establishing a stable and dependable cloud detection algorithm stands as a crucial component in the preprocessing of optical satellite data [6]. Traditional cloud detection techniques predominantly rely on threshold-based methods that classify individual pixels based on their spectral properties [4]. These methods operate by setting predefined threshold values for spectral reflectance in specific bands, such as: Visible (VIS) - Clouds often appear bright with high reflectance, Near-Infrared (NIR) - Clouds typically exhibit strong reflectance in this band, Shortwave Infrared (SWIR) - This band helps differentiate between clouds, snow, and certain land features. This improved version offers better flow, clarity, and sentence structure. While these methods are simple and computationally efficient, they face notable limitations. One such limitation is the Lack of Spatial Context, where Threshold methods assess each pixel independently, ignoring spatial relationships between neighbouring pixels [4]. This limitation can lead to misclassifications in areas with similar spectral properties. Surfaces such as snow, ice, or urban structures often resemble cloud characteristics in spectral data, leading to frequent false positives [8]. Cloud properties vary widely in density, altitude, and optical thickness. Static threshold values often fail to effectively account for this variability. Factors such as sun angle, seasonal changes, or atmospheric effects can alter reflectance values, reducing the reliability of fixed threshold criteria [23]. This revision improves clarity, organisation, and readability while maintaining technical precision [1].

Present threshold methods primarily hinge on categorising spectral attributes of individual pixels in isolation, lacking integration of spatial information. Consequently, this frequently results in misclassifications, particularly on luminous surfaces like artificial structures or snow/ice. While multi-temporal techniques offer some relief, securing cloud-free imagery of the scene poses a formidable challenge. The research aims to identify the unique spectral and spatial features of cloud-covered

areas by extracting characteristics that distinguish cloud pixels from non-cloud pixels using multispectral data. It also leverages information from the entire image patch rather than analysing pixels in isolation, thereby improving robustness by capturing spatial relationships crucial for accurate cloud detection [2]. Multiple convolution layers are employed to progressively extract low-level features (e.g., textures and edges) and high-level features (e.g., shapes and patterns) in a bottom-up manner. Interpolation methods are used to ensure the seamless recovery of missing data points in partially obscured imagery, thereby enhancing spatial coherence. Additionally, the research emphasises the development of orientation-sensitive features to improve alignment precision in cloud segmentation. Orientation-invariant features further ensure consistent performance under varying lighting conditions and cloud patterns.

The study emphasises the development of an interpretable cloud segmentation method using feature semantic learning, enhancing the model's transparency in decision-making. The focus on Landsat-8 multispectral imagery ensures a practical and scalable solution for real-world applications. The adoption of a Deep Convolutional Neural Network (DCNN) with an enhanced encoder-decoder architecture aims to improve feature extraction and segmentation precision. The model is rigorously tested on Landsat-8 datasets, highlighting its robustness in challenging conditions such as thin clouds, snow-covered regions, and mixed land types. The following is the research project's goal: 1. To determine the area's feature expression where the detected object belongs. 2. To utilise the complete picture patch's contextual information. To use several convolution modules to extract features from the bottom up. 3. To use pixel-wise interpolation to rebuild the complete feature map. 4. Creating orientation-sensitive and orientation-invariant features is essential to reducing the discrepancy between classification scores and localisation accuracy. The goal is to utilise feature-semantic learning to develop an interpretable cloud segmentation technique. For reliable cloud detection, concentrate on Landsat-8 multispectral photography.

Implement an improved encoder-decoder architecture for a deep convolutional neural network (DCNN). Analyse model performance using Landsat-8 datasets, with a particular focus on the model's resilience in challenging situations. Simple Linear Iterative Clustering (SLIC) is a widely used image segmentation algorithm that clusters pixels based on colour similarity and spatial proximity. In this research, the SLIC method is employed to partition the image into superpixel sub-regions, which are groups of pixels with homogeneous properties. Within these superpixels, Bright and dense cloud superpixels are easily identifiable due to their high reflectance. Clearly, non-cloud superpixels (e.g., water bodies, vegetation) are also distinguishable. This step enhances computational efficiency by narrowing down regions that require more complex analysis. A brightness-based thresholding function is applied to distinguish Obvious cloud superpixels with strong reflectance, Non-cloud regions with lower reflectance values, and Potential cloud regions that require further examination. This approach accelerates the detection process by avoiding unnecessary analysis of clearly identifiable regions. Fully Convolutional Neural Networks (FCNs) are deep learning models designed for pixel-level classification tasks [10]. FCNs utilise: Convolution layers to extract multi-scale features, and Up-sampling layers to generate pixel-wise prediction maps. FCNs excel in capturing both spectral and spatial information, improving cloud detection accuracy in complex environments.

The Fuzzy C-Means (FCM) clustering algorithm partitions image data into multiple fuzzy clusters, assigning degrees of membership to each pixel [3]. This method is particularly effective in handling ambiguous regions such as cloud edges, semi-transparent clouds, or mixed pixels [15]. Additionally, a modified cloud shadow index calculation further refines the detection process by excluding false positives, such as pseudo-cloud shadows caused by variations in land cover [5]. The research work initially employs the Simple Linear Iterative Clustering (SLIC) method to partition the colour composite image into superpixel sub-regions. Within these super pixels, bright and dense cloud super pixels, as well as clearly non-cloud super pixels, are readily discerned. Subsequently, a threshold function based on brightness features is applied to differentiate between obvious cloud and non-cloud super pixels, as well as those potentially containing clouds, thereby expediting the detection process. Additionally, the research explores the application of fully convolutional neural networks for cloud detection, highlighting their ability to leverage spectral and spatial information in image classification tasks with sufficient training data. Moreover, it explores the application of the fuzzy c-means method (FCM) for cloud and cloud shadow detection, along with a modified calculation of the cloud shadow index to exclude certain pseudo cloud shadows.

2. Literature Review

In this study, Wang et al. [18] propose a cloud intrusion detection system that utilises a stacked contractive autoencoder (SCAE) for feature reduction and a support vector machine (SVM) for classification, aiming to enhance detection performance in cloud environments. While effective in handling large and redundant network traffic, the method faces challenges of complexity and maintenance. It outperforms other state-of-the-art methods on the KDD Cup 99 and NSL-KDD datasets. Guo et al. [7] present ClouDet, a lightweight deep-learning framework for cloud detection in remote sensing imagery. Utilising a dilated separable convolutional module, ClouDet efficiently extracts multiscale contextual information, offering large receptive fields with fewer parameters and lower computational complexity. The framework incorporates a fusion strategy of multiple features and context pooling to improve accuracy and effectiveness, achieving superior results on several public cloud detection datasets compared to state-of-the-art methods. ClouDet is particularly well-suited for embedded platforms due to its minimal time and memory

requirements, as well as its high precision. Irish et al. [15] introduce CDL, a cloud detection algorithm for land areas, utilising the microwave humidity sounder-2 (MWHS-2) and a gradient boosting decision tree (GBDT) for enhanced detection efficiency. The algorithm leverages data from China's CINRAD weather radar to improve accuracy and generalisation, using various frequency groups from MWHS-2. The temperature channel near 118.75 GHz (tem-algorithm) exhibits superior performance compared to other algorithms operating in different frequency ranges. CDL demonstrates high detection accuracy and reduced computational cost, but is less suitable for large-scale scenarios. It is effective for quality control and cloud filtering in atmospheric and surface parameter retrievals. Using the fuzzy c-means (FCM) algorithm, Bo et al. [13] created a technique for detecting clouds and cloud shadows in satellite photos. This method reduces misclassification by utilising a modified cloud shadow index computation and enhances detection accuracy by more accurately characterising pixel states, particularly for thin clouds. It is validated using 41 Landsat photos of different types of land cover and requires fewer preset parameters. The FCM method demonstrates more accurate detection than the support vector machine (SVM) approach and fewer false positives than the function of mask (Fmask) method.

In their analysis of cloud systems, intrusion detection, and blockchain applications for counteracting cyberattacks, Alkadi et al. [12] place a strong emphasis on collaborative anomaly detection for identifying both insider and external threats. The study examines how blockchain technology might improve trust management and data privacy in cloud-based Network Intrusion Detection Systems (NIDS). It highlights the challenges and future directions for integrating these technologies, noting the need for research in live migration processes and attack detection. Blockchain emerges as a promising solution for ensuring trust and security in cloud operations. Trepte et al. [14] discuss the improvements in the CERES Edition 4 (Ed4) cloud detection system, which utilises Terra and Aqua MODIS data. Ed4 enhances cloud detection accuracy by utilising additional spectral channels, revised calibrations, and new models, achieving better performance than its predecessor, Ed2, particularly in non-polar regions [16]. The Ed4 system shows high accuracy in identifying cloudy or clear areas, with consistency between Aqua and Terra data, although challenges remain in polar regions and over land. These advancements help CERES create a more accurate radiation budget and contribute to a cloud property climate data record. Sun et al. [9] introduce the Surface Reflectance-Based Cloud Shadow Detection (SRCSD) algorithm for Landsat 8 OLI images, leveraging prior knowledge of background land surface reflectance to improve cloud shadow detection accuracy.

The SRCSD algorithm uses MODIS-derived surface reflectance data to estimate top-of-atmosphere reflectance and identify shadow-covered pixels, demonstrating higher accuracy and efficiency compared to the Fmask method, particularly in detecting thin and broken cloud shadows [20]. The algorithm shows promise for adaptation to various satellite data types, though its computational burden may limit its application in real scenarios. Validation results show producer and user accuracy rates of approximately 0.805 and 0.893, respectively. Bo et al. [13] present the Object-oriented Cloud and Cloud-shadow Matching (OCM) method for cloud detection in Chinese moderate to high-resolution satellite imagery, utilising a modified Automatic Cloud Cover Assessment (ACCA) approach to generate initial cloud and shadow maps [21]. The OCM method effectively pairs cloud and cloud-shadow objects to improve detection accuracy, achieving an overall accuracy of nearly 90% across 200 HJ-1/CCD and GF-1/WFV images in China. Despite challenges like the absence of SWIR and TIR bands, the method demonstrates superior accuracy compared to other state-of-the-art techniques. However, it requires a significant amount of time and resources to construct. King et al. [10] introduce three primary techniques for cloud detection using MODIS data [24]. The first method involves spectral threshold tests, which exploit MODIS's diverse spectral bands to distinguish clouds based on their reflectance properties in visible (VIS), near-infrared (NIR), and thermal infrared (TIR) wavelengths. The second technique applies a spatial coherence test that identifies clouds by examining pixel consistency and detecting characteristic variability patterns typical of cloud formations.

The third approach employs a temporal consistency test, which monitors cloud behaviour across consecutive satellite observations to enhance detection reliability and reduce false positives. By integrating these strategies, the research effectively improves cloud detection accuracy under varying environmental and geographical conditions. Irish et al. [15] introduced multiple strategies for detecting clouds using historical geostationary satellite sensors for climate research. One method involves applying multi-spectral thresholding, which examines variations in reflectance and temperature across different spectral bands to identify cloud presence [16]. Another technique incorporates temporal consistency checks, ensuring stable cloud detection by observing patterns over time and minimising false alarms. Additionally, the research employs spatial context analysis, where relationships between neighbouring pixels are used to enhance detection accuracy, particularly in complex scenarios such as thin clouds or diverse land types. By integrating these methods, the study significantly enhances cloud detection capabilities for long-term climate observation. Liu et al. [16] present various strategies for accurately calibrating the CLARREO infrared spectrometer to support climate change detection. One key approach involves defining strict spectral calibration standards to ensure high radiometric precision across different wavelengths. The research highlights the use of onboard calibration references, which provide stable benchmarks to minimise instrument drift and ensure long-term measurement reliability. Moreover, the study presents a comprehensive error analysis framework to assess uncertainties in spectral data, thereby enhancing the robustness of climate observations.

These combined methods effectively improve the accuracy and dependability of climate data obtained from the CLARREO infrared spectrometer. Chen et al. [22] presented the Iterative Haze-Optimised Transformation (IHOT) method, which aims to automatically detect clouds and haze in Landsat imagery. This technique uses an iterative optimisation process that gradually enhances the distinction between clear-sky pixels, haze, and cloud regions. The method employs spectral feature analysis, with a particular focus on the blue and shortwave infrared (SWIR) bands, to effectively identify haze-impacted areas from clouds. Furthermore, the IHOT algorithm integrates edge-preserving methods to ensure that region boundaries remain sharp during the detection process. By combining iterative refinement, spectral analysis, and boundary-preserving mechanisms, this method significantly boosts the precision and dependability of cloud and haze detection in satellite images. Sun et al. [9] introduce a cloud detection method in their work that leverages visible to shortwave infrared (SWIR) data for enhanced precision [17]. The approach integrates thresholding techniques with spectral signature analysis to effectively distinguish clouds from non-cloud areas [22]. By examining reflectance patterns across particular wavelength bands, the method improves cloud identification under various lighting and environmental conditions. Furthermore, the algorithm utilises adaptive thresholding to account for variations in surface properties and atmospheric effects. This well-rounded approach ensures reliable cloud detection in diverse geographical settings and complex weather environments.

Drusch et al. [11] introduce a cloud detection technique that integrates Support Vector Machine (SVM) classification with the SWIR2 spectral band and tasselled cap transformation to improve accuracy. The SWIR2 band proves useful in differentiating clouds from reflective surfaces, such as snow or sand, due to its sensitivity to moisture levels. By applying the tasselled cap transformation, key attributes such as brightness, greenness, and wetness are extracted, enhancing the identification of cloud regions [17]. The SVM classifier is trained using these extracted features, enabling it to effectively distinguish cloud pixels from non-cloud ones. This integrated method significantly improves cloud detection capabilities, particularly in environments with complex surface characteristics. Bao et al. [20] present a swift cloud detection technique that utilises the low-frequency elements of satellite images. The method involves converting satellite data into the frequency domain through processes such as the Discrete Cosine Transform (DCT) or the Discrete Wavelet Transform (DWT). By concentrating on low-frequency components, the approach effectively highlights large cloud formations while reducing the impact of high-frequency noise and fine details. A thresholding method is then employed to distinguish clouded regions from clear areas using these low-frequency traits. This technique enables fast and reliable cloud detection, making it ideal for large-scale or real-time remote sensing tasks [19]. Chen et al. [22] present a method for detecting clouds and cloud shadows using a multiscale 3D Convolutional Neural Network (3D-CNN) tailored for high-resolution multispectral imagery.

This technique combines spectral and spatial data by utilising 3D convolution layers, which process information across multiple spectral bands and spatial dimensions simultaneously [19]. By adopting a multiscale framework, the method effectively captures features at varying resolutions, enhancing its ability to identify clouds with different shapes and sizes. Moreover, the 3D-CNN model excels at distinguishing between clouds, cloud shadows, and land cover by leveraging contextual details. This approach significantly improves detection accuracy, especially in complex scenarios involving challenging cloud patterns. Fenta et al. [1] developed a deep learning method designed to achieve accurate and dependable cloud detection in satellite images. CloudFCN leverages a Fully Convolutional Network (FCN), which analyses entire image segments rather than individual pixels, thereby improving both processing speed and context awareness [24]. The model undergoes training on a variety of satellite image datasets, ensuring its adaptability across different environmental and atmospheric settings. Through the use of multiple convolutional layers, the method efficiently captures spatial patterns, enhancing its ability to differentiate clouds from land surfaces. CloudFCN proves to be highly effective, particularly in challenging conditions involving thin clouds or complex backgrounds. Zhu and Woodcock [24] present Cloud-Net, a specialised algorithm designed for cloud detection in Landsat 8 imagery. This method utilises a deep convolutional neural network (CNN) that efficiently performs end-to-end cloud detection, eliminating the need for significant pre- or post-processing [25]. The network is designed to capture both spatial and spectral information, enhancing its ability to accurately identify clouds in diverse environmental conditions. The model is trained using extensive Landsat 8 data, ensuring reliable performance across various atmospheric scenarios. Cloud-Net effectively detects both dense and thin clouds, establishing itself as a robust tool for cloud identification in satellite images.

3. Case and Methodology

The proposed system presents a modified version of the U-Net architecture, specifically designed for cloud identification in multispectral remote sensing images. This architecture only uses two down-sampling steps, in contrast to the original U-Net. There are also fewer filters for each convolutional layer, and depth-wise separable convolutions are used instead of traditional convolutions, drastically reducing the number of trainable parameters. The input used for training and testing is top-of-atmosphere (TOA) calibrated radiance. To maximise network training, TOA reflectance inputs are further normalised using the mean and standard deviation from the L8-Biome dataset for each band. To minimise the pixel-wise binary cross-entropy between model predictions and ground truth labels, the networks are trained from scratch. In this study, we use a modified version of the U-Net architecture. Our U-Net architectural modifications are designed to accelerate training and testing, while providing a simplified framework that can serve as a basis for future research studies. Interestingly, compared to the original

U-Net's five down-sampling steps, our suggested architecture only has two. To drastically reduce the trainable parameters, we have also reduced the number of filters for each convolutional layer and replaced traditional convolutions with depth-wise separable convolutions.

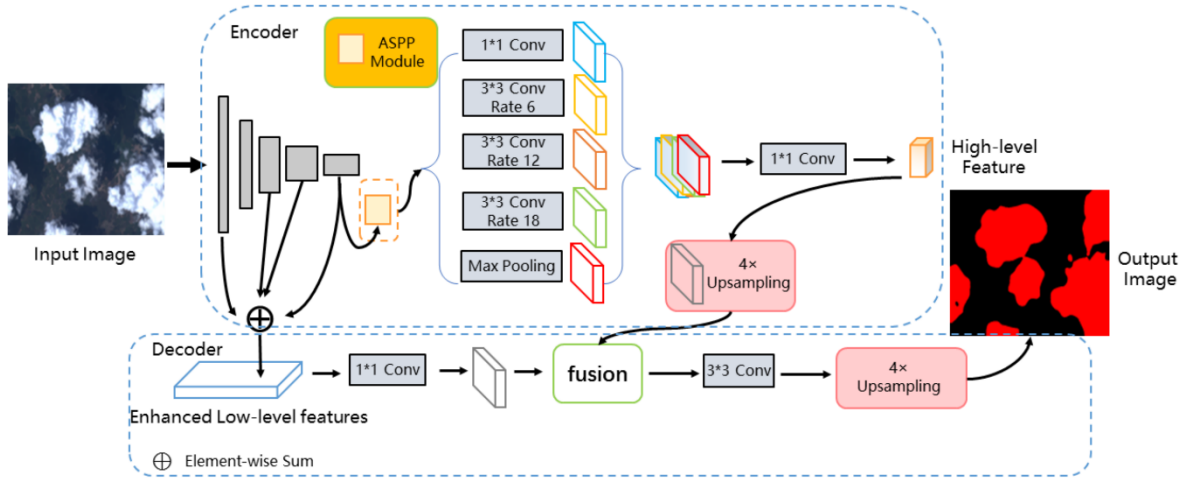

Figure 1: System architecture of the proposed ED-CNN [26]

Figure 1 illustrates the project's architecture diagram. Firstly, the input is obtained from the camera, and the algorithm is trained using datasets that serve as training data. The input from the camera will extract the captured images. The images are then forwarded to the SSD model, and the detection model is loaded using MobileNetV2.Then, image processing techniques are applied to the captured image, and mask classifiers determine whether the person is wearing a mask by outlining the face area. Our entire work is divided into three modules. The preprocessing module primarily focuses on splitting the train and test data. There is no image pre-processing since atrous spatial pooling is used to improve model interpretability. The data augmentation modules enhance the training dataset by creating new datapoints through image transformation techniques. The final module is where the crux of the model is implemented using an enhanced encoder-decoder architecture.

3.1. Dataset Description

This collection includes the carefully derived pixel-level ground facts for cloud detection from 38 Landsat 8 scene photos. Although it is a variation of the dataset, the 38-Cloud dataset is introduced in the document. For deep learning-based semantic segmentation algorithms, the full photos of these sceneries are cropped into several 384*384 patches. There are 9201 testing patches and 8400 training patches. Near Infrared (band 5), Blue (band 2), Green (band 3), and Red (band 4) are the four spectral channels that match each patch. These channels are not integrated, unlike other computer vision images. For deep learning-based semantic segmentation algorithms, the full photos of these sceneries are cropped into several 384*384 patches. Cloud detection is a crucial step in many optical imagery-based remote sensing applications. The 95-Cloud dataset is a comprehensive dataset designed to assist researchers in assessing their cloud segmentation algorithms that rely on deep learning [27].

This set comprises 34,701 384x384 training patches. The exam set for 95-Cloud is identical to that of 38-Cloud. Seventy-five Landsat 8 Collection 1 Level-1 scenes, primarily in North America, are used to extract the training patches. The test set for 95-Cloud consists of 9,201 patches, each with 20 scenes. The 95-Cloud and 38-Cloud test sets are identical. However, their training regimens differ. Compared to 38-Cloud, 95-Cloud has 57 additional training scenes. Each patch includes four related spectral channels: Near Infrared (band 5), Blue (band 2), Green (band 3), and Red (band 4). These channels are not integrated, unlike other computer vision images. Rather, they are located in the appropriate directories. The model's performance is usually tested using this dataset. Both the 38-cloud and 95-cloud datasets were employed in our paper to assess the performance of our model. Given that the model was trained on the Landsat-8 dataset and assessed on the Sentinel-2 dataset, its accuracy of 98.7% was expected. The availability of additional datasets, such as SPARCS, could significantly enhance the generalisation of models.

3.2. Pre-Processing

Reflectance values in Landsat 8 imagery were adjusted to top-of-atmosphere reflectance before processing. Except for the 15 m panchromatic band (B8), all spectral bands were utilised. Nine predictive characteristics were derived by combining the two thermal bands to create a single thermal feature. This was done to prevent any false information from arising from different processing methods related to the stray light problem and to facilitate the network's adaptation to other Landsat sensors. The

training dataset's feature-wise mean and standard deviation were used to normalise each feature. Seventy-two subscenes were reserved for training after six were set aside for validation and two for testing. A 64-pixel border of no data was added to these 1000px × 1000px subscenes to mimic predictions at the margins of the Landsat scenes. Each 256px × 256px window requested during training was randomly selected from one of these padded subscenes. The network was permitted to continue for up to 100 epochs or until the validation sample ceased improving for five epochs, whichever came first. Each epoch has 1440 samples from all subscenes. In practice, the absence of validation improvement caused all epochs to finish early. In accordance with established practices, training examples were randomised and batched. The network uses a 256 px × 256 px window for training and predicts a central 200 px × 200 px region.

3.3. Data Augmentation

For supervised semantic segmentation, image augmentation is crucial when the number of labels is limited. In this context, the act of creating additional training samples by transforming existing annotated images in various ways is referred to as data augmentation. The performance and generalisation of the supervised semantic segmentation model are improved by this method, which successfully expands the variety and volume of available training data. Data augmentation is important because it can mitigate the impact of label scarcity. There are often few annotated images available, as gathering precise pixel-level annotations for semantic segmentation can be expensive and time-consuming. However, fresh samples can be produced using two primary ways by supplementing the current labelled data. (i) It is possible to use image distortion and corruption techniques such as rotation, noise addition, and pixel distortion.

3.4. Feature Semantic Learning Network

Two-dimensional (2D) convolution layers are interspersed with 2 * 2 max pooling layers to represent the convolution phase. As the max pooling layers reduce the effective resolution, N (and hence the depth of the subsequent convolution) increases. Each 2D convolutional layer learns N filters of size 3 × 3 × Depth. Resolution is decreased by the max pooling layers, which analyse every non-overlapping 2 × 2 window and only pass through the maximum value. A growing number of filters that describe more intricate spatial relationships are traded for resolution as a result of this process, which aggregates spatial information throughout the image. Approximately 16.5 million of the network's 20.5 million weights are present in the suggested design at this point, matching the architecture used in transfer learning. The encoder utilises convolutional layers for multi-scale feature extraction, as shown in Equation 1. To enhance low-level features, the algorithm combines the upsampled high-level features with the low-level ones using element-wise summation, as shown in Equation 2.

$$F_{conv} = \sigma(W * I + b) \quad (1)$$

Where:

- F_{conv} = Output Feature map
- W = Convolution Filter
- I = Input Feature Map
- b = Bias Term
- σ = ReLU Activation Function

The spatial resolution is reconstructed during the deconvolution phase using information that encodes the spatial structure. In this stage, 2D deconvolution layers—also known as the transpose of 2D convolution are utilised.

$$F_{fusion} = F_{low} + \text{Upsample}(F_{high}) \quad (2)$$

Where:

- F_{low} = low-level Features
- F_{high} = High-level Features
- $Upsample$ = Upsampling operation to match spatial dimensions

The decoder refines the fused features using convolution and up-sampling as in Equation 3, and the final cloud mask is derived as in Equation 4.

$$F_{decoder} = \sigma(W_{dec} * F_{fusion}) \quad (3)$$

$$M_{cloud} = \begin{cases} 1 & \text{if } F_{decoder}(x,y) > T \\ 0 & \text{Otherwise} \end{cases} \quad (4)$$

Where:

M_{cloud} = Final binary cloud mask

T = Threshold value

These layers have four distinct outputs grouped in a 2 × 2 window, and they learn filters of size 2 × 2 × Depth. Each deconvolution layer's output doubles the input's resolution in this manner. In contrast to the convolution phase, fewer filters are learnt and employed for prediction as the resolution rises. The network may utilise both the spatial information reconstructed from the large-scale features and the more moderate-scale data, as the moderate-resolution features from Phase 1 are directly incorporated into the deconvolution outputs following two of these upscaling stages. The data is entirely restored to its initial resolution during the deconvolution process. Our network's output phase has a unique feature: data flow divides into two branches to deter the network from relying solely on fine-scaled features. In the first branch, no fine-scale spatial features are provided; thus, the network must acquire relevant characteristics for the classification job during the convolution and deconvolution stages. However, both branches predict the same labels. The second allows the network to fine-tune spatial structure by combining the output from the deconvolution phase with the early fine-scale characteristics. The aggregate loss from these outputs is weighted twice as heavily as the loss from the second. Prediction only utilises the second branch, which contains the fine-scale features. Furthermore, a regularisation layer called 2D spatial dropout is employed. The network is forced to acquire redundant patterns that ideally match distinct lines of evidence because three out of eight characteristics are set to 0 during training. The dropout is left out during prediction so that every characteristic can aid in classification.

4. Result and Discussion

Two separate input datasets were utilised for this paper. The Landsat-8 and the Sentinel-2 datasets. Both datasets capture satellite cloud images of various types over North America. Each dataset consisted of the same number of data points. The Landsat-8 dataset was used for the training process, with the ratios of datapoints belonging to each of the four labels (gravel, sugar, flower, and fish) set at 1:1:1:1. The Sentinel-2 dataset was used for the validation and testing processes, ensuring the model generalised well to unseen data. The Landsat-8 satellite images featuring clouds of various types are shown in Figure 2.

Figure 2: Landsat-8 satellite images featuring clouds of various types

Figure 3 below showcases the model performance on identifying and classifying clouds belonging to the following four mesoscale cloud patterns:

- **Sugar:** very fine-scale cloud dusting with minimal vertical extension and minimal signs of self-organisation (due to gust fronts or cold pools). Sugar got its name because the clouds that formed resembled a dusting of powdered sugar.
- **Gravel:** There is no indication of associated stratiform cloud veils, but the cloud fields are structured along meso-β (20 to 100 km) lines or arcs that define cells with intermediate granularity and brighter cloud elements (relative to

Sugar). Both the brightness contrast and the granularity of the cloud-defined patterns were higher in Gravel than in Sugar. More significantly, Gravel clouds were arranged in arcs or lines, believed to be connected to gust fronts accompanying cold pools. Brighter, presumably deeper clouds were seen to demarcate the areas where gust fronts clashed, where new cells frequently formed. Gravel occasionally displayed features that resembled open mesoscale cellular convection.

- **Fish:** Meso-β scale stratiform cloud features with irregular shapes, ranging from 20 to 200 km. They frequently have greater reflectivity cores and appear in quasi-regularly spaced bunches, hence the plural, with individual features clearly separated from one another by cloud-free zones. The most unexpected and unique system of organisation was seen in flowers. They consist of patches of stratiform clouds at the meso-β scale, frequently accompanied by centre clusters that are embedded and support the stratiform cloud patches. The “Flowers” pattern's individual flowers (or stratiform patches) range in size from a few tens to a few hundred kilometres in diameter.
- **Flowers:** Irregularly shaped meso-β scale (20 to 200 km) stratiform cloud features, often with higher reflectivity cores, and appearing in quasi-regularly spaced bunches (hence the plural) with individual features well separated from one another by regions devoid of clouds. Flowers were the most surprising and distinct pattern of organisation. They are comprised of meso-β scale patches of stratiform clouds, often with evidence of central clusters embedded and supporting these patches. The scale of an individual Flower (or stratiform patch) in the pattern “Flowers” varies from a few tens to a few hundreds of kilometres.

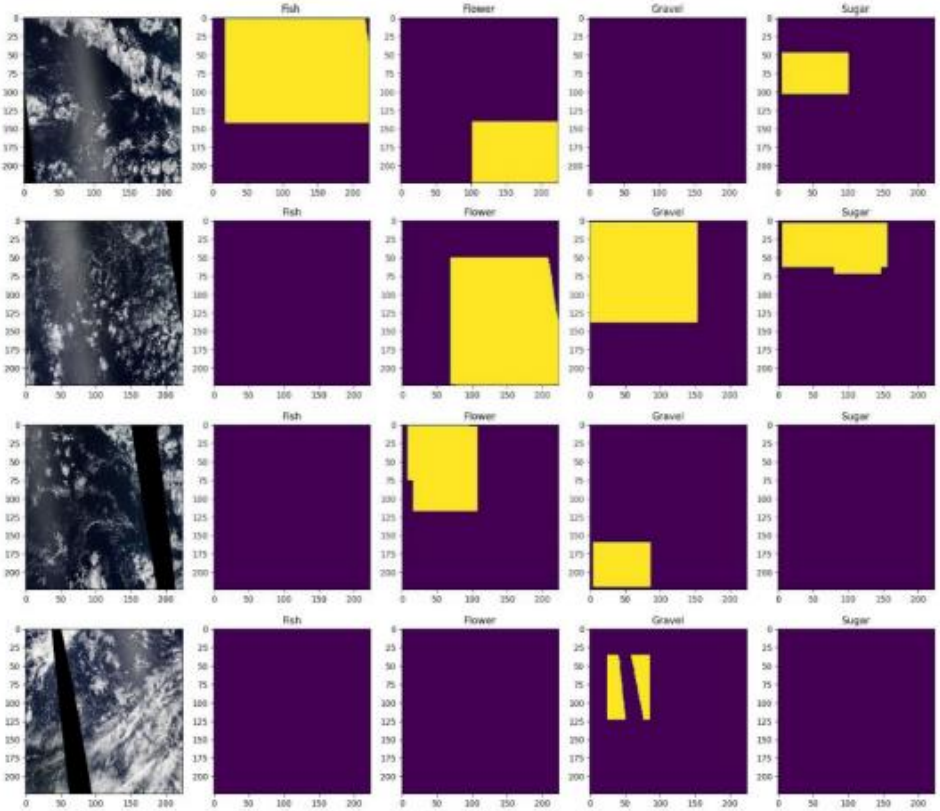

Figure 3: Cloud segmentation through feature semantic learning

The proposed system enhances the existing system by simplifying the model, enhancing the efficiency of extracting and merging multi-scale features, and reducing the computational resources required. It also outperforms the existing system in terms of accuracy, efficiency, and the ability to detect thin and broken cloud shadows. While the existing system relies on manual labelling of a large number of samples, the proposed system utilises unsupervised training, thereby reducing the time and cost associated with data annotation. To evaluate the model's performance and potential for overfitting, the training loss and validation loss were monitored throughout the training process. Figure 4 depicts the loss curves for all shape types. The training loss generally exhibits a decreasing trend across epochs for all shapes, indicating the model's ability to learn and improve its representation of the training data. The validation loss ideally follows a similar trend, suggesting the model generalises well to unseen data. This suggests that the model achieves a good balance between fitting the training data and generalising to unseen data. Additionally, the training process employs a validation-based model selection strategy, whereby

models with the lowest training loss are saved only if they also achieve an acceptable validation loss. This approach further helps mitigate overfitting and select the best-performing model for each shape type.

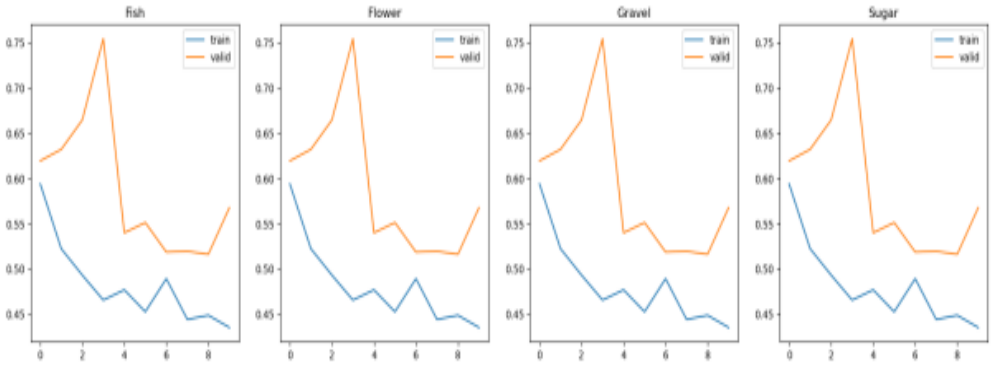

Figure 4: Measuring loss through n=10 iterations

On each iteration, the Landsat-8 dataset was utilised for the training process, meaning it determined the model's parameters and influenced its adjustments. The Sentinel-2 dataset was then utilised to test the model classifier's performance on unseen data. An acceptable loss value, initially set to 0.5, is determined based on whether the model classifier was saved. If a new epoch performed better than the acceptable loss, the loss value was updated to reflect this improvement, and the model was rewritten. We continued this process for an initially decided iteration number of 10. However, this number is at the user's discretion. Upon completion of the validation process, an average accuracy percentage of 97.5% was recorded across all model shape types. This outperforms several existing models, including CNN (90.92%), M-ResNet (93.36%), and SSCA-net (96.91%).

5. Conclusion

Cloud detection for remote sensing images is a critical procedure in this study because clouds are common in optical remote sensing images, making several tasks, including target recognition, environmental monitoring, and land cover monitoring, extremely difficult. This research proposes a new technique for detecting clouds in multispectral remote sensing photos from Landsat 8. The SLIC method is initially used to cluster the colour composite image into superpixel sub-regions. In the colour composite image of Bands 6, 3, and 2, clouds are clearly visible and can be easily separated from snow and ice. The thick and brilliant cloud super pixels are clearly distinguished from the non-cloud super pixels. To expedite the identification process, we utilise a threshold function in the brightness feature space to distinguish probable cloud super pixels from obvious cloud and non-cloud super pixels. This paper presents cloud detection models based on recent advancements in deep learning. In particular, we choose fully convolutional neural networks, which excel at exploiting spectral and spatial information in image classification problems when enough training data are available to learn the model parameters. We have two primary objectives for our efforts in this regard. On the one hand, we wish to investigate the potential for model transfer between different satellite sensors. When labelled data are unavailable for a particular sensor, this would lessen the need for training data. Consequently, we aim to develop a precise cloud detection model that can be utilised with Sentinel, which is directly learned from actual Landsat 2 images.

5.1. Future Work

To increase the precision of our cloud detection technique, we will explore additional data in the future. We'll take the cloud shadow, for instance, as it always appears in pairs with clouds. Furthermore, we will modify our technique to simultaneously identify water, ice/snow, cloud shadows, and both thick and thin clouds. A fully held-out test set, not used for training or validation, can be used to assess the model's performance and further investigate its generalizability. This provides a more reliable assessment of the model's performance on unseen data. A more thorough investigation into using data augmentation techniques to artificially expand the quantity and variety of training data may also be necessary. By doing so, some problems with data imbalance may be resolved, and the model's capacity to generalise to hidden changes in the shapes may be enhanced. Additional improvements include the model's ability to function consistently across several satellite bands.

Acknowledgement: The authors from SRM Institute of Science and Technology, Sathyabama Institute of Science and Technology, and Saveetha Engineering College sincerely acknowledge the support of their institutions and all participants for their meaningful contributions to the success of this research.

Data Availability Statement: The study utilises a dataset on interpretable and pattern-aware cloud segmentation using feature semantic learning. All data used in this research are available and properly documented for reference.

Funding Statement: This study and manuscript were conducted by the authors without any external financial assistance.

Conflicts of Interest Statement: The authors confirm that they have no conflicts of interest concerning this study, and all sources, citations, and references have been accurately acknowledged.

Ethics and Consent Statement: The authors confirm that the relevant organisations granted ethical approval, and informed consent was obtained from all participants, ensuring full compliance with ethical research standards.

References

1. A. A. Fenta, H. Yasuda, N. Haregeweyn, A. S. Belay, Z. Hadush, M. A. Gebremedhin, and G. Mekonnen, "The dynamics of urban expansion and land use/land cover changes using remote sensing and spatial metrics: The case of Mekelle City of northern Ethiopia," *International Journal of Remote Sensing*, vol. 38, no. 14, pp. 4107–4129, 2017.
2. A. Yang, B. Zhong, W. Lv, S. Wu, and Q. Liu, "Cross-calibration of GF-1/WFV over a desert site using Landsat-8/OLI imagery and ZY-3/TLC data," *Remote Sensing*, vol. 7, no. 8, pp. 10763–10787, 2015.
3. B. Zhong, A. Yang, A. Nie, Y. Yao, H. Zhang, and S. Wu, "Finer resolution land-cover mapping using multiple classifiers and multisource remotely sensed data in the Heihe River Basin," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 10, pp. 4973–4992, 2015.
4. B. Zhong, P. Ma, A. Nie, A. Yang, Y. Yanjuan, L. Wenbo, Z. Hang, and Q. Liu, "Land cover mapping using time series HJ-1/CCD data," *Science China Earth Sciences*, vol. 57, no. 8, pp. 1790–1799, 2014.
5. B. Zhong, W. Chen, S. Wu, L. Hu, Q. Liu, and X. Luo, "A cloud detection method based on relationship between objects of cloud and cloud-shadow for Chinese moderate to high resolution satellite imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 11, pp. 4898–4908, 2017.
6. B. Zhong, Y. Zhang, T. Du, A. Yang, W. Lv, and Q. Liu, "Cross-calibration of HJ-1/CCD over a desert site using Landsat ETM+ imagery and ASTER GDEM product," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 7247–7263, 2014.
7. H. Guo, H. Bai, and W. Qin, "ClouDet: A dilated separable CNN-based cloud detection framework for remote sensing imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, no. 9, pp. 9743–9755, 2021.
8. K. Hu, D. Zhang, M. Xia, M. Qian, and B. Chen, "LCDNet: Light-weighted cloud detection network for high-resolution remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, no. 6, pp. 4809–4823, 2022.
9. L. Sun, Q. Wang, X. Zhou, J. Wei, X. Yang, W. Zhang, and N. Ma, "A priori surface reflectance-based cloud shadow detection algorithm for Landsat 8 OLI," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 10, pp. 1610–1614, 2018.
10. M. D. King, S. Platnick, W. P. Menzel, S. A. Ackerman, and P. A. Hubanks, "Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua satellites," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 7, pp. 3826–3852, 2013.
11. M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini, "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sensing of Environment*, vol. 120, no. 5, pp. 25–36, 2012.
12. O. Alkadi, N. Moustafa, and B. Turnbull, "A review of intrusion detection and blockchain applications in the cloud: Approaches, challenges and solutions," *IEEE Access*, vol. 8, no. 6, pp. 104893–104917, 2020.
13. P. Bo, F. Su, and Y. Meng, "A cloud and cloud shadow detection method based on fuzzy c-means algorithm," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, no. 4, pp. 1714–1727, 2020.
14. Q. Z. Trepte, P. Minnis, S. Sun-Mack, C. R. Yost, Y. Chen, and Z. Jin, "Global cloud detection for CERES Edition 4 using Terra and Aqua MODIS data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9410–9449, 2019.
15. R. R. Irish, J. L. Barker, S. N. Goward, and T. Arvidson, "Characterization of the Landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm," *Photogrammetric Engineering & Remote Sensing*, vol. 72, no. 10, pp. 1179–1188, 2006.
16. S. Liu, Y. Yin, Z. Chu, and S. An, "CDL: A cloud detection algorithm over land for MWHS-2 based on the gradient boosting decision tree," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, no. 8, pp. 4542–4549, 2020.

17. T. Wu, X. Hu, Y. Zhang, L. Zhang, P. Tao, and L. Lu, “Automatic cloud detection for high resolution satellite stereo images and its application in terrain extraction,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 121, no. 11, pp. 143–156, 2016.
18. W. Wang, X. Du, D. Shan, R. Qin, and N. Wang, “Cloud intrusion detection method based on stacked contractive auto-encoder and support vector machine,” *IEEE Transactions on Cloud Computing*, vol. 10, no. 3, pp. 1634–1646, 2022.
19. X. Y. Zhuge, X. Zou, and Y. Wang, “A fast cloud detection algorithm applicable to monitoring and nowcasting of daytime cloud systems,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 11, pp. 6111–6119, 2017.
20. Y. Bao, L. Liu, and J. Wang, “Estimating biophysical and biochemical parameters and yield of winter wheat based on Landsat TM images,” *in Proc. IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*, Boston, MA, United States of America, 2008.
21. Y. Chen, L. Tang, W. Huang, J. Guo, and G. Yang, “A novel spectral indices-driven spectral–spatial-context attention network for automatic cloud detection,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, no. 3, pp. 3092–3103, 2023.
22. Y. Chen, Q. Weng, L. Tang, Q. Liu, X. Zhang, and M. Bilal, “Automatic mapping of urban green spaces using a geospatial neural network,” *GIScience & Remote Sensing*, vol. 58, no. 4, pp. 624–642, 2021.
23. Y. Chen, Q. Weng, L. Tang, X. Zhang, M. Bilal, and Q. Li, “Thick clouds removing from multitemporal Landsat images using spatiotemporal neural networks,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, no. 12, pp. 1–14, 2020.
24. Z. Zhu and C. E. Woodcock, “Object-based cloud and cloud shadow detection in Landsat imagery,” *Remote Sensing of Environment*, vol. 118, no. 3, pp. 83–94, 2012.
25. Z. Zhu, S. Wang, and C. E. Woodcock, “Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel-2 images,” *Remote Sensing of Environment*, vol. 159, no. 1, pp. 269–277, 2015.
26. K. Zheng, J. Li, L. Ding, J. Yang, X. Zhang, and X. Zhang, "Cloud and Snow Segmentation in Satellite Images Using an Encoder–Decoder Deep Convolutional Neural Networks," *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 7, p. 462, 2021, doi: 10.3390/ijgi10070462.
27. S. Mohajerani, 38-Cloud-A-Cloud-Segmentation-Dataset: This data set includes Landsat 8 images and their manually extracted pixel-level ground truths for cloud detection. Available Online: Dataset URL: https://github.com/SorourMo/38-Cloud-A-Cloud-Segmentation-Dataset [Accessed by 01/02/2024].