



Hybrid Deep and Machine Learning Framework for Cloud and Shadow Segmentation in Landsat-8 Imagery

Isro Tri Pambudi*, Sani Muhamad Isa

Computer Science Department, BINUS Graduate Program –Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia

*isro.pambudi@binus.ac.id

Abstract. Cloud and shadow interference in satellite imagery reduces the quality and reliability of remote sensing data. The traditional method would face issue to predict data near the shadow and cloud. To address this challenge, this study is focus improve the accuracy the area near shadow and cloud detection in Landsat-8 imagery. The implementation of hybrid module using standard CNN and U-Net CNN and a machine learning model using K-Nearest Neighbors (KNN) on SPARCS and CCA18 Landsat 8 dataset. A hybrid approach was then implemented by integrating CNN outputs and metadata into the second model (KNN/RF), and final evaluation was conducted using accuracy metrics. The research results show that the proposed hybrid deep and machine learning approach improves the accuracy of cloud and shadow segmentation in Landsat-8 imagery. Additionally, the implementation demonstrates that this method can reduce manual effort and computational cost, making it suitable for researchers with limited resources.

Keywords: Cascading, Hybrid model, Landsat-8 SPARCS, Remote Sensing, Semantic Segmentation

(Received 2025-05-10, Revised 2025-07-12, Accepted 2025-07-17, Available Online by 2025-10-30)

1. Introduction

Satellite imagery refers to images produced by satellites and is commonly used as a basis for decision-making. The utilization of high-resolution satellite imagery can be applied across various fields [1]. The process that researchers must perform before utilizing satellite imagery begins with a procedure known as masking. Masking is necessary to ensure that the obtained information is free from noise [2]. Noise refers to data that provide biased information; in the case of satellite imagery, this is caused by clouds obscuring the objects of interest or cloud shadows that alter the values from their true representation. The process can be time-consuming; therefore, additional technologies such as machine learning and deep learning are needed to assist researchers.

Supervised machine learning, which involves gathering features such as spectrum, intensity, color temperature, camera correction, and geometric characteristics, can be challenging to collect [3]. In the study conducted by He, the use of RGB spectrum made the process relatively simple compared to

Landsat 8 data. To address this issue, unsupervised machine learning is employed. Among the various deep learning models, the most complex algorithm is the Convolutional Neural Network (CNN) [4]. The implementation of Convolutional Neural Network (CNN) technology is not new and has been commonly used by previous researchers, such as Hughes. CNN has been implemented to detect several target objects, including clear-sky, clouds, shadows, and snow [2]. Based on the research conducted by Kennedy, the accuracy for each object varies, with the highest accuracy reaching 98.1% and the lowest at 90.5%. The other implementation is using CNN-based Cloud Masking (CCM), but the accuracy below 85%, an need a significant improvement [5].

To improve the accuracy of cloud and shadow segmentation, the principle of solar direction determination can be utilized. The determination of shadow and cloud shadow direction can be achieved through the calculation of the Apparent Solar Azimuth [6]. Based on Ibrahim's research, the cloud direction was obtained with an angular accuracy range of less than 15 degrees. The calculation process of the apparent solar azimuth is formulated in Eq. (1) below, where φ_s and θ_s represent the solar azimuth and zenith angles, respectively. The angle formation includes two types of angles based on the difference in radians.

$$\tan \tan (\varphi_a) = (\sin \varphi_s \tan \theta_s - \sin \varphi_v \tan \theta_v) / (\cos \varphi_s \cos \theta_s - \cos \varphi_v \cos \theta_v) \quad (1)$$

The accuracy of prediction results can be improved through post-processing. This can be achieved by utilizing shadow direction information found in the dataset metadata and performing further predictions using the K-Nearest Neighbor (KNN) method. The implementation of KNN offers advantages in pixel-based classification [7]. In this study, a hybrid model is implemented by initially performing predictions using a CNN and KNN. The results of predictions are then enhanced with additional parameters for further predictions using a subsequent model. This method has been implemented and has a better result [8]. The implementation of machine learning can be used multiple times and with the additional of the data the better model can be generated, the process in machine learning also can be used repeatedly based on the usage case [9].

Previous study by [10] state that the Learning Attention Network Algorithm (LANA), a deep learning-based method for cloud and cloud shadow detection in Landsat imagery that outperforms traditional methods like Fmask and the U-Net Wieland model. Unlike previous pixel-based approaches, LANA leverages spatial attention within a U-Net architecture to enhance detection, particularly for challenging classes such as thin clouds and shadows. Trained on a large and diverse annotated dataset, LANA achieved higher classification accuracy (up to 88.84%) and better F1-scores across all classes. It also demonstrated superior performance in temporal smoothness analysis, indicating fewer undetected clouds and shadows. The study emphasizes the importance of high-quality annotations, identifies challenges in generalizing across sensors, and suggests LANA's potential for broader application, including for future sensor data like Sentinel-2 and MSS, provided proper retraining is done.

[11] Also state that a comprehensive overview of cloud detection algorithms in optical remote sensing, emphasizing deep learning-based approaches. It outlines the limitations of traditional methods and demonstrates how models such as FCN, encoder–decoder, attention mechanisms, and GAN offer improved accuracy and adaptability. The analysis compares these methods to classical and machine learning techniques, highlighting their respective advantages and drawbacks. It also examines public datasets and essential post-processing steps like cloud shadow detection and removal. Ultimately, the survey identifies current challenges and suggests future directions to enhance the effectiveness and precision of deep learning models in cloud detection tasks.

This study introduces a novel hybrid framework that combines deep learning and machine learning methods for cloud and cloud shadow segmentation in Landsat-8 imagery. Unlike previous approaches that rely solely on one technique, this research integrates the spatial-contextual strengths of deep learning with the stable and interpretable classification capabilities of machine learning. The primary objective is to enhance the accuracy of detecting complex features such as thin clouds and subtle cloud shadows, while maintaining consistent performance across varying imaging conditions. By leveraging

the complementary strengths of both approaches, the study aims to deliver a more robust and reliable segmentation solution to support advanced remote sensing analysis.

2. Literature Review

2.1 Landsat-8 Spectral Bands

Landsat is well known satellite for generate imagery and has been started from Landsat-1, the usage of the the imagery satellite is give a huge resources to a lot of the development of the Landsat satellite is continue and the newest one is Landsat-9 [12]. Landsat-8 is the satellite that being used to capture images. The Landsat-8 has been launched since 2013 with the implementation to be operated for 5 years [13], but it still operated until now. it still operated. The features used in the development of the CNN model are derived from the information in the spectral bands. The sensor that being used in Landsat-8 has implemented a system to minimize the effect of atmospheric particles to ensure the data quality [14]. The detailed explanation of each Landsat 8 spectral band based on information from NASA are: (NASA Goddard Space Flight Center)

- (i) Band 1 (Coastal/Aerosol): Wavelength range of 0.435 - 0.451 μm , used for detecting aerosol particles in the atmosphere, such as in coastal mapping activities.
- (ii) Band 2 (Blue): Wavelength range of 0.452 - 0.512 μm , useful for mapping deep water zones and coastal areas. Additionally, it can be used for vegetation mapping.
- (iii) Band 3 (Green): Wavelength range of 0.533 - 0.590 μm , applicable for vegetation and water analysis.
- (iv) Band 4 (Red): Wavelength range of 0.636 - 0.673 μm , used for vegetation analysis or land classification.
- (v) Band 5 (NIR - Near-Infrared): Wavelength range of 0.851 - 0.879 μm , suitable for analyzing vegetation biomass and vegetation changes.
- (vi) Band 6 (SWIR-1 - Short-Wave Infrared 1): Wavelength range of 1.566 - 1.651 μm , capable of penetrating clouds to extract information beneath them. It can also be used for soil moisture and vegetation analysis.
- (vii) Band 7 (SWIR-2 - Short-Wave Infrared 2): Wavelength range of 2.107 - 2.294 μm , useful for studying mineral rocks and vegetation.
- (viii) Band 8 (Pan - Panchromatic): Wavelength range of 0.503 - 0.676 μm , provides higher spatial resolution and more detailed satellite imagery.
- (ix) Band 9 (Cirrus): Wavelength range of 1.363 - 1.384 μm , specifically designed to detect cirrus clouds, which are high-altitude, thin clouds that can affect land observation.
- (x) Band 10 (TIR-1 - Thermal Infrared 1): Wavelength range of 10.60 – 11.19 μm , used for measuring Earth's surface temperature and emitted heat.
- (xi) Band 11 (TIR-2 - Thermal Infrared 2): Wavelength range of 11.50 – 12.51 μm , provides additional data for measuring surface temperature.

2.2 Convolutional Neural Network

Convolutional Neural Network (CNN) is an advancement of neural networks and represents one of the most efficient versions (Figure 1). CNN is an algorithm that establishes connections between data, mimicking the structure of the human brain. It features a multi-layer neural network, enabling the model to learn and extract features automatically without human intervention. CNN is commonly used in developing models for tasks such as image segmentation and natural language processing, particularly in 2D data formats [15].

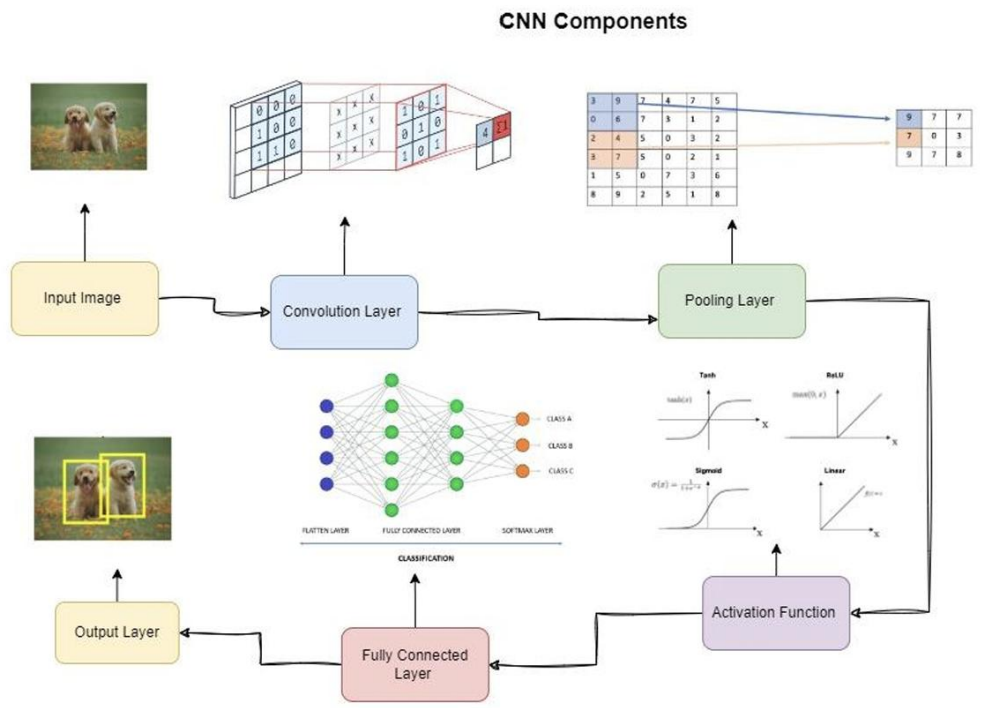


Figure 1. CNN Components [15]

Convolutional Neural Network (CNN) has been effective for several usage of image satellite data such as land use [16]. CNN model can be used for supervised learning for the classification from the available data that has been through labelling process [17].

2.3 K-Nearest Neighbour

K-Nearest Neighbor (KNN) is a machine learning algorithm that can be used for classification and regression tasks. KNN is a non-parametric method that identifies similarities based on the K nearest samples [7]. This algorithm considers spatial relationships to determine the class or value based on the closest K samples. KNN is widely used in pattern recognition, image processing, and multimedia data analysis due to its ease of implementation and high accuracy in many cases. In the context of remote sensing, predictions can be modeled based on properties and shapes, making KNN a suitable approach for classifying data.

2.4 Random Forest

Random Forest (Random Forest) is a machine learning algorithm that well-known for the classification data. It implement ensemble of decision tree that randomly selected based on the data given [18]. Random forest work as parallel classifier, that means on the training process every single decision tree would be through training independently. The difference of the parallel process would make the ensemble method can be implemented.

2.5 Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is the machine learning algorithm that really good for classify large and complex dataset [18], the usage of the XGBoost can be used to handle data noise and occurrence of the outliers data. The algorithm works well in detecting patern based on important rating. The scability and the speed of XGBoost works well for the big data such as Landsat-8.

2.6 Cloud and Shadow Association

To detect the relationship between buildings and their shadows, it was concluded that buildings generate

shadows when illuminated by sunlight, with the shadows resembling the shape and distance of the buildings [19]. This occurs because light cannot fully penetrate solid structures, causing shadows to form as light particles are blocked by the buildings (Figure 2). The study also explains that the sun's angle relative to the building significantly influences the shape and distance of the resulting shadow. When buildings are closely packed, the shadows tend to overlap and partially cover adjacent structures.

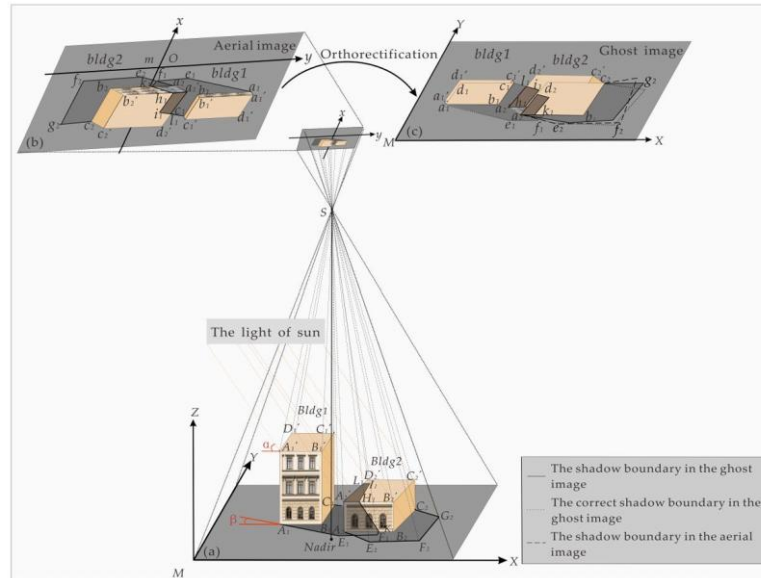


Figure 2. Building and shadow association [19].

2.7 Ensemble and Cascading Method

In this study, a hybrid model is implemented by initially performing predictions using a machine learning model. The prediction results are then enriched with additional parameters for use in subsequent predictions. Cascading and ensemble methods can be applied to improve accuracy, which is referred to as the Improved Neural Network method [8]. The Improved Neural Network method incorporates both ensemble and cascading approaches. The implementation of ensemble combined with cascading has been proven to enhance the accuracy of the developed model. According to the study, the cascading method specifically demonstrated superior accuracy compared to other implementations.

3. Methods

3.1 Data Identification

The Landsat 8 data was obtained through the USGS webpage using the Landsat-8 SPARCS dataset. This dataset has been previously utilized for CNN model to predict data in remote sensing [20] (Figure 3). The distribution of satellite images is evenly spread across almost all continents. Metadata information is available for each location, which includes details about the solar direction.

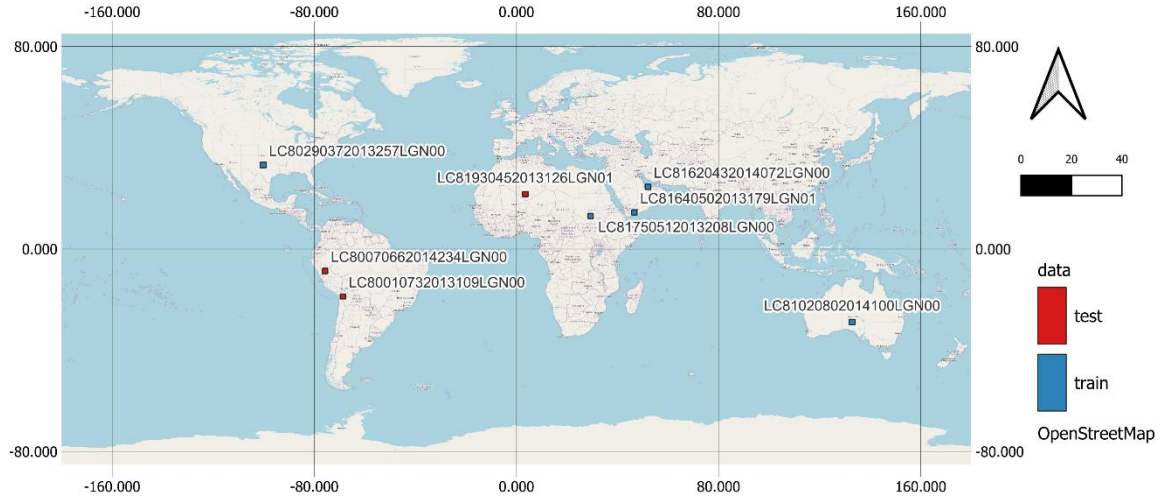


Figure 3. Data distribution

The dataset contains the distribution of interpretation results into six main categories. This information was manually created by analysts, with a total of 8 satellite images, 5 dataset will be treated as data train and the 3 dataset will be treated as data test. These images are provided in TIF file format, which retains spatial information within the image. Additionally, the metadata includes information related to solar azimuth, which is crucial as a primary feature for developing the KNN model. The distribution of prediction results (Table 1).

Table 1. The distribution of prediction results.

Category	Total	Percentage
Fill	16,959,968	26.12%
Shadow	1,390,173	2.14%
Thin Cloud	37,061,292	57.06%
Cloud	1,104,429	1.70%
Other	1,936,429	2.98%

3.2 Data Preprocessing

Preprocessing is carried out through several main steps, including identifying missing values, incorrect data types, handle different bands range, and handling outlier values. In cases of missing values or incorrect data types, the missing data can be filtered out. Data cleaning can also be done using standard scaling; however, this is part of the optimization process. This ensures that the resulting model is not overly rigid and can still adapt to other data cases [21]. One of the process of data preprocessing for Landsat-8 data atmospheric correction, it will make sure the value of the raw band is clear from the atmosphere altered value (NASA Goddard Space Flight Center).

The preprocessing stage in this study is essential to ensure the Landsat-8 imagery is suitable for input into the hybrid deep and machine learning framework. The process involves several key steps, starting with the identification and handling of missing values and incorrect data types, which are either filtered out or imputed using statistical techniques such as mean substitution. Next, normalization is applied to account for differences in the range of spectral bands, typically using standard scaling to ensure uniformity across input features. Outlier values, which may distort the learning process, are identified using methods such as Z-score or interquartile range (IQR) filtering and subsequently removed. A critical part of preprocessing specific to satellite imagery is atmospheric correction, which converts raw digital numbers into surface reflectance by eliminating atmospheric effects such as scattering and

absorption. This correction ensures that the reflectance values more accurately represent true surface conditions. Together, these preprocessing steps not only improve data quality but also enhance the robustness and adaptability of the model across various environmental conditions.

3.3 Data Splitting

Based on the cleaned data from the previous stage, the dataset will be split 5 dataset will be used as the training data meanwhile the other 3 dataset be used as testing data. For the training The process for generate hybrid model the dataset will be split into 80% data training and 20% data validation for CNN. Meanwhile for KNN or RF Model will be used all train data for training. This separation is performed using packages provided by scikit-learn. An important point to consider is the treatment process for the data, such as applying functions like median, mode, and mean, which need to be conducted separately for each dataset. The separation between training and final testing dataset is used to make sure the result can be applicable for another image satellite and make the model become more reliable. Each of dataset created on different location, time, sun position, and angle of the satellite camera.

3.4 Convolutional Neural Network Model Creation

The model development in this study is divided into two main stages: the CNN model and the KNN model. The first CNN model uses an $11 \times 11 \times 16$ input as the initial input. In the first convolutional pooling stage, data reduction is performed because Hughes research disregarded padding at the edges of the image. The convolutional layer employs the ReLU activation function. The architecture also incorporates BatchNorm2d to accelerate the process and downsampling, which is carried out using MaxPool2d in the second convolutional pooling stage. This approach is necessary due to the large size of satellite imagery (Table 2).

Table 2. The parameters used for building the CNN model.

Block	Layer	Parameter [size, stride] or weight	Size
input	Reshape	-	$1 \times 10 \times 1$
CP1	Conv2D	[3×3, stride=1]	$1 \times 10 \times 32$
	ReLU	-	$1 \times 10 \times 32$
	MaxPool2D	[2×2, stride=2]	$1 \times 5 \times 32$
CP2	Conv2D	[3×3, stride=1]	$1 \times 5 \times 64$
	ReLU	-	$1 \times 5 \times 64$
	MaxPool2D	[2×2, stride=2]	$1 \times 2 \times 64$
FC1	Flatten	-	128
	Dense	128 units	128
	Dropout	rate=0.5	128
FC2	Dense (Output)	3 units (softmax)	3

The second CNN model used is CNN Unet model. This model is the advances model that can boost the model accuracy. The implementation of CNN Unet using encoder and decoder which can boost the accuracy (Pang, et. al., 2023). The model that has been generated will be more complex than a simple CNN model (Table 3).

Table 3. The parameters used for building the CNN model.

Block	Layer	Parameter	Size
Input	Input Layer	-	$256 \times 256 \times 11$
CP1	Conv2D	64 filters, 3×3, stride=1	$256 \times 256 \times 64$

CP2	BatchNorm2D	-	$256 \times 256 \times 64$
	MaxPooling2D	2×2 , stride=2	$128 \times 128 \times 64$
	Conv2D	128 filters, 3×3 , stride=1	$128 \times 128 \times 128$
	BatchNorm2D	-	$128 \times 128 \times 128$
CP3	MaxPooling2D	2×2 , stride=2	$64 \times 64 \times 128$
	Conv2D	256 filters, 3×3 , stride=1	$64 \times 64 \times 256$
	BatchNorm2D	-	$64 \times 64 \times 256$
	MaxPooling2D	2×2 , stride=2	$32 \times 32 \times 256$
CP4	Conv2D	512 filters, 3×3 , stride=1	$32 \times 32 \times 512$
	BatchNorm2D	-	$32 \times 32 \times 512$
	MaxPooling2D	2×2 , stride=2	$16 \times 16 \times 512$
	MaxPooling2D	2×2 , stride=2	$16 \times 16 \times 512$
Bottleneck	Conv2D	1024 filters, 3×3 , stride=1	$16 \times 16 \times 1024$
Decoder1	Conv2DTranspose	512 filters, 2×2 , stride=2	$32 \times 32 \times 512$
	Concatenate	0	$32 \times 32 \times 1024$
Decoder2	Conv2D	512 filters, 3×3 , stride=1	$32 \times 32 \times 512$
	Conv2DTranspose	256 filters, 2×2 , stride=2	$64 \times 64 \times 256$
	Concatenate	0	$64 \times 64 \times 512$
	Conv2D	256 filters, 3×3 , stride=1	$64 \times 64 \times 256$
Decoder3	Conv2DTranspose	128 filters, 2×2 , stride=2	$128 \times 128 \times 128$
	Concatenate	0	$128 \times 128 \times 256$
	Conv2D	128 filters, 3×3 , stride=1	$128 \times 128 \times 128$
	Conv2DTranspose	64 filters, 2×2 , stride=2	$256 \times 256 \times 64$
Decoder4	Concatenate	0	$256 \times 256 \times 128$
	Conv2D	64 filters, 3×3 , stride=1	$256 \times 256 \times 64$
	Conv2D	64 filters, 3×3 , stride=1	$256 \times 256 \times 64$
	Conv2D	64 filters, 3×3 , stride=1	$256 \times 256 \times 64$
Output	Conv2D	5 filters (softmax), 1×1	$256 \times 256 \times 5$

3.5 The second model for hybrid model

The second model that being used to create hybrid model is K Nearest Neighbor (KNN). KNN model is one of the model that can be used to predict image satellite data [7]. The implementation of KNN can be used to do the classification automatically [22]. For the specific case the accuracy of the KNN model can reach 94% [23]. The second model that being implemented for the comparison for KNN model is Random Forest (RF), because RF has been implemented in remote sensing for the spatial data to target distribution of the forest fire [24]. The usage of Random Forest is used for the benchmarking the efficiency of the accuracy for the second model prediction for the hybrid model. The development of the second model involves creating features based on the prediction results of the CNN model, specifically the accuracy of cloud, shadow, and other classifications. Additional features, such as sun direction and spatial location, are also included to enhance the second model. The construction and training of the second model incorporate experimental thresholds set at 80%, 70%, and 60%. These thresholds are applied to ensure that predictions are not performed on data with accuracy exceeding the specified values (Figure 4). This approach helps to optimize the model's performance and focus computational resources on areas requiring further refinement.

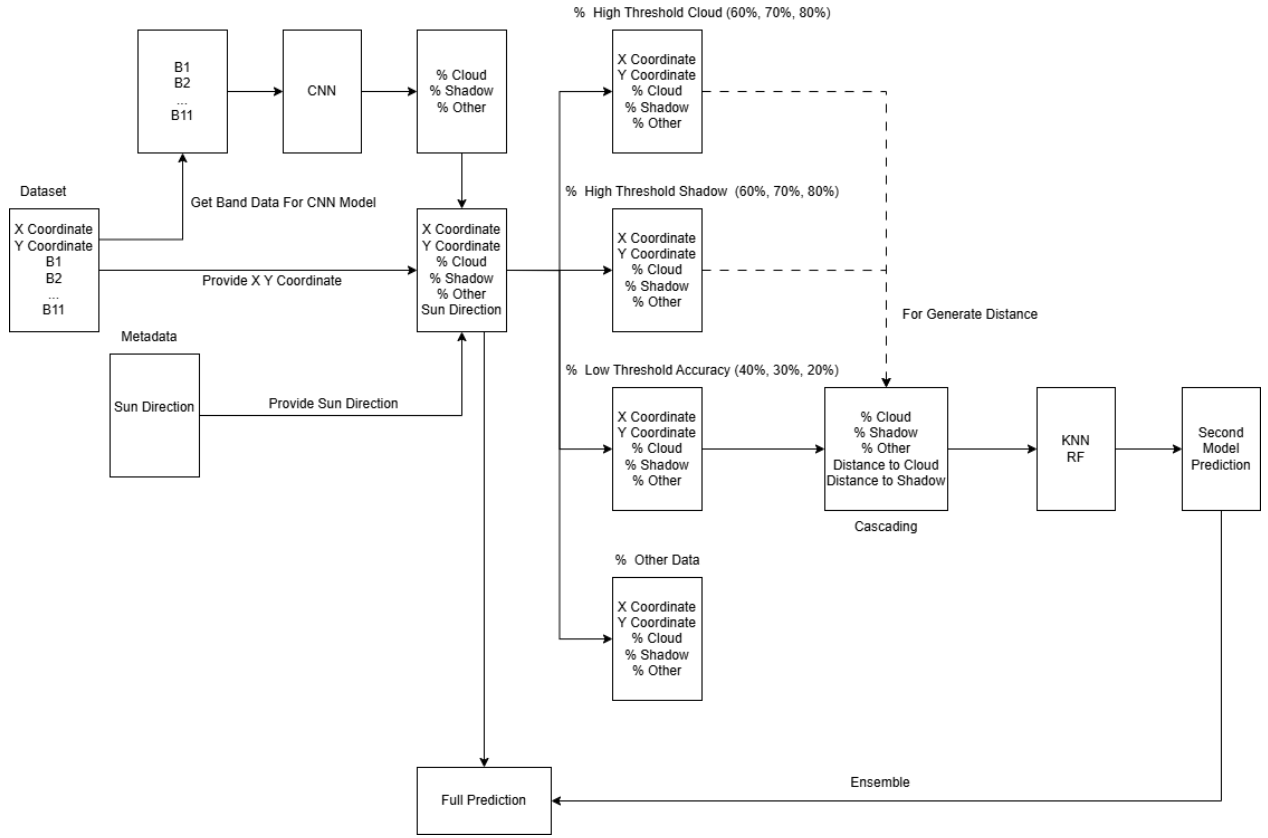


Figure 4. The Flow of Hybrid Concept

3.6 Model Evaluation

In the Keras library, there are methods to evaluate a model and obtain performance metrics such as accuracy Eq. (2). These evaluation parameters are reliably used for neural network models, particularly in Convolutional Neural Networks (CNN) cases [25]. Accuracy is the proportion of correct predictions made by the model. In addition, the model evaluation will include an analysis of the categorical cross-entropy loss graph during the training process using CNN. This graph provides insights into the model's performance and learning behaviour over time.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

4. Results and Discussion

4.1. Data Preparation

The dataset used in this study, as shown in Table 4, comprises eight scenes five designated for training and three for testing. Previously utilized by researchers such as [20], this dataset includes imagery from various regions captured under different conditions and timeframes. The diversity of the data is essential for developing a robust model capable of generalizing across varying scenarios; however, it also presents significant challenges in the preprocessing stage.

Table 4. USGS Raster to generate the model.

Stage	Biome	Landsat ID	Cloud Cover
Training	Barren	LC81640502013179LGN01	7.33%
	Grass / Crops	LC80290372013257LGN00	25.61%
	Grass / Crops	LC81750512013208LGN00	58.97%
	Shrubland	LC81020802014100LGN00	76.28%
	Urban	LC81620432014072LGN00	20.29%
Test	Barren	LC81930452013126LGN01	64.06%
	Forest	LC80070662014234LGN00	5.05%
	Shrubland	LC80010732013109LGN00	6.71%

Based on this research, the publicly accessible dataset from USGS must undergo a preprocessing stage that includes atmospheric correction and normalization. Several tools, such as ENVI, Google Earth Engine (GEE), and GRASS, can assist in performing atmospheric correction. In this study, a simplified atmospheric correction was conducted following the guidelines provided in the Landsat 8 (L8) Data Users Handbook. Although the results of this correction are limited, they are sufficient to produce a more reliable dataset from the original imagery. Normalization was applied using standard scaling, transforming the data into a range between 0 and 1, as shown in Table 5.

Table 5. Preprocessing Landsat 8 Images LC81640502013179LGN01

Bands	Raw					Atmospheric Correction and Normalization				
	Count	Min	Max	Mean	Std	Count	Min	Max	Mean	Std
Band-1	58452291	0	45,372	9,803.84	6,546.13	58,452,291	0.0	1.0	0.216	0.121
Band-2	58452291	0	46,948	9,798.42	6,577.34	58,452,291	0.0	1.0	0.208	0.116
Band-3	58452291	0	46,778	10,488.91	7,068.56	58,452,291	0.0	1.0	0.224	0.135
Band-4	58452291	0	48,910	11,968.76	8,172.4	58,452,291	0.0	1.0	0.244	0.152
Band-5	58452291	0	52,262	13,944.82	9,625.84	58,452,291	0.0	1.0	0.266	0.185
Band-6	58452291	0	65,535	15,962.56	11,100.83	58,452,291	0.0	1.0	0.243	0.161
Band-7	58452291	0	62,125	14,194.41	9,797.64	58,452,291	0.0	1.0	0.228	0.142
Band-8	58452291	0	47,154	10,969.65	7,405.04	58,452,291	0.0	1.0	0.232	0.143
Band-9	58452291	0	9,681	3,692.18	2,377.81	58,452,291	0.0	1.0	0.381	0.190
Band-10	58452291	0	45,092	23,087.15	15,692.14	58,452,291	0.0	1.0	0.512	0.303
Band-11	58452291	0	38,080	21,022.65	14,238.26	58,452,291	0.0	1.0	0.552	0.378

Previous researchers applied transfer learning on different datasets to improve model accuracy [20]. Although this study also uses Landsat-8 imagery, the clean dataset employed here differs significantly, as it was processed using a simpler atmospheric correction method. As a result, the model's accuracy may be slightly lower compared to studies that used more refined preprocessing. Additionally, normalization was not fully optimized due to limited resources for atmospheric correction. However, this research focuses on utilizing cloud and shadow direction as added features to enhance CNN model performance, and the current dataset remains sufficient to achieve this objective.

4.2. Training and Testing Model with CNN and CNN Unet

The initial CNN model in this research was developed based on the approach by [2], which demonstrated strong performance using the SPARCS Landsat-8 dataset. Model training was conducted on Google Colab T4, utilizing 50 GB of RAM and a GPU with up to 16 GB. Raster images were divided into 256×256 patches to build the model, with a maximum batch size of 4 due to hardware limitations. Initially, training with raw data resulted in poor performance, achieving only 32.97% accuracy, indicating the

necessity of a proper preprocessing step. After applying atmospheric correction and normalization, the model's accuracy improved to 53.65%. Further improvements through hyperparameter tuning—specifically, increasing epochs and reducing the learning rate—raised the accuracy to 71.32%. Finally, modifying the architecture to use a CNN U-Net model significantly boosted accuracy to 81.23%, as shown in Table 6.

Table 6. Comparison of Overall Accuracy based on previous researcher.

Sequence Method	Overall Accuracy
CNN Model, Predict using raw data (batch size = 4, epoch = 5, learning rate=0.001)	32.97%
CNN Model , Atmospheric Correction with Normalization (batch size = 4, epoch = 5, learning rate=0.001)	53.65%
CNN Model (batch size = 4, epoch = 6, learning rate=0.00001)	71.32%
CNN Unet Model (batch size = 4, epoch = 6, learning rate=0.00001)	81.27%

Hyperparameter tuning, model architecture, and dataset quality play a crucial role in determining the performance of the generated model. Adjusting the learning rate, in particular, has shown a significant impact on accuracy; in this study, a lower learning rate yielded better model performance. Given that the dataset consists of over 50 million raster pixels, a smaller learning rate is beneficial for capturing complex patterns and achieving more stable convergence. Although increasing the number of epochs showed improvement, the difference between epoch 5 and epoch 6 was minimal, likely due to the large and diverse dataset already providing sufficient learning opportunities in earlier iterations. Previous studies, such as those by [26] and [27], also highlight the importance of tuning learning rates and model structures, especially when dealing with high-resolution satellite imagery, to maximize segmentation performance.

4.3. Comparison CNN and CNN Unet with the previous researcher

The comparison of the model accuracy with the previous researcher is required to make sure that this research will give a better understanding of the quality of model that generated and implementation of the hybrid model later on. There are some previous researchers published papers using the same raw dataset, but every researchers used different method and additional information to make sure the model generation will be great. On the generation a good model additional dataset would give a greater result, and actually some of the researchers using transfer learning add other datasets that successfully boosted the accuracy of the model (Table 7).

Table 7. Comparison of Overall Accuracy based on previous researcher.

Model	Overall Accuracy
Zhu's method	87.71%
Li's method	88.95%
Zi's method	91.16%
CNN	71.32%
CNN Unet	81.27%

Zi's method demonstrates a notable advantage due to its use of a two-step superpixel classification strategy, combining a double PCA Network with a Support Vector Machine to enhance classifier performance. In contrast, this research focuses on utilizing cloud and shadow directional features to improve model accuracy. Although the CNN U-Net model in this study achieved an accuracy of 81.27%, which may not be ideal for direct comparison with Zi's approach, it remains valid within the study's context. Zi's findings also acknowledge the importance of incorporating cloud and shadow

relationships, which supports the premise of this research. Moreover, given the limited computational resources and the absence of an external dataset for training, the performance of the CNN U-Net model is considered acceptable, especially as the primary objective is to evaluate the impact of cloud and shadow distance features on enhancing model prediction accuracy.

4.3. CNN Unet Result Analysis

This initial result of CNN Unet from testing dataset generate a different accuracy from the each of the testing scene. The range of the accuracy show a quite large difference of accuracy. The better result comes from model LC81930452013126LGN01, and the lowest one is LC80010732013109LGN00. The accuracy differences come from a several factors the first factor is the different dataset statistic value, the second factor come from the difference of Biome, the third factors come from the different location and position, and the fourth factor is the imbalance data (Table 7).

Table 7. CNN Unet detail prediction.

Model	LC81930452013126LGN0	LC80070662014234LGN0	LC80010732013109LGN0
	1	0	0
Accuracy	87.22%	82.41%	74.19%

The initial raw dataset exhibits varying statistical characteristics, even after atmospheric correction is applied, with noticeable differences in minimum and mean values between training and testing datasets. These discrepancies significantly affect model accuracy, particularly because low minimum values are often interpreted as "Fill" data, which can distort model predictions (Table 8). For example, Band 1, which captures atmospheric aerosol particles, displays varying mean values, indicating different data distributions—especially in raster LC80010732013109LGN00, which belongs to a barren biome with a low cloud cover (7.33%) that is vastly different from the over-30% cloud cover seen in the training data.

This mismatch negatively impacts model performance. Conversely, rasters like LC80070662014234LGN00 and LC81930452013126LGN01, which are located in forest and shrubland biomes respectively, show better accuracy due to biome similarity with training data (which includes grassland and shrubland). This indicates that similarity in biome and cloud cover between training and testing datasets leads to improved model accuracy. Additionally, the dataset is highly imbalanced, with Cloud, Shadow, and other categories each comprising less than 10% of the total data, which poses further challenges for model training and classification.

Table 8. Basic Statistics of Atmospheric Correction and Normalized of Band 1 in dataset.

Scene	Type	Min	Max	Mean	Std
LC81640502013179LGN01	Train	0.231156	1.0	0.304418	0.049333
LC80290372013257LGN00		0.152812	1.0	0.209326	0.064557
LC81750512013208LGN00		0.151095	1.0	0.313551	0.161350
LC81020802014100LGN00		0.183382	1.0	0.441112	0.140887
LC81620432014072LGN00		0.167775	1.0	0.282354	0.087249
LC81930452013126LGN01	Test	0.154284	1.0	0.401218	0.184777
LC80070662014234LGN00		0.151295	1.0	0.200698	0.054720
LC80010732013109LGN00		0.116258	1.0	0.206243	0.093744

4.4. Implementation of Cascading

The implementation of cascading is a crucial step in developing the hybrid model, as it enables the dataset to be enriched with two additional features: the distance to cloud and the distance to shadow. These distance features are derived from the softmax output of the initial CNN model, which provides class probabilities for cloud, thin cloud, shadow, fill, and other. Not all softmax outputs are used in the

second-stage model; instead, thresholds are applied to selectively filter the data. Two types of thresholds are introduced: the top threshold, which identifies high-confidence predictions (e.g., 60% to 90%) for temporary classification of cloud or shadow based on softmax scores, and the bottom threshold (set at 50% and 40%) to capture uncertain predictions requiring further analysis. With the sun azimuth angle as a reference, distances to cloud and shadow objects are calculated and added to the dataset. This approach ensures that only ambiguous or low-confidence classifications (below the bottom threshold) are passed to the second model for refinement, improving overall prediction accuracy through informed data enhancement.

Table 9. Threshold vs the second dataset for the next model

Top Threshold	Bottom Threshold	All	Retrain	Percentage
90%	40%	58,452,291	418,817	0.71651084
80%		58,452,291	543,136	0.9291954
70%		58,452,291	632,953	1.08285405
60%		58,452,291	749,184	1.28170169
90%	50%	58,452,291	1,551,976	2.65511578
80%		58,452,291	3,914,029	6.6961088
70%		58,452,291	5,243,476	8.97052264
60%		58,452,291	6,593,840	11.2807212

4.5. Comparison of the second model

The retrain of the cascading will be train with the second model with three different model which are KNN (K Nearest Neighbor), RF (Random Forest), and XGBoost. The usage of three model will be used as a comparison to make sure the concept of adding shadow and cloud distance have an effect for the accuracy of the model prediction. The point that would be focus is the changes in percentage of the retrain dataset from the CNN prediction and the second model prediction. Based on the dataset and the architecture model, the higher value of accuracy would give implications of the shadow and cloud distance to predict the cloud and shadow. The positive result on the result also tell that the research have correctly implement the cascading method for the dataset and the model creation. The result of comparison result can be shown on (Table 10).

Table 10. Threshold vs the second dataset for the next model

Model	Top Threshold Percentage	Bottom Threshold Percentage	Increase Accuracy From CNN Percentage
KNN	90	40	9.188546731
	80		4.956111359
	70		2.381118637
	60		2.638317978
RF	90		6.554294976
	80		2.64100475
	70		1.876032865
	60		3.111741082
XGBoost	90	50	8.304700162
	80		5.86155478
	70		5.283616135
	60		4.349633829
KNN	90	50	-2.692840065

	80	2.498011976
	70	-2.175121468
	60	-2.295652544
	90	-3.710156711
RF	80	0.863693034
	70	-3.251195312
	60	-2.401452448
	90	-2.362768727
XGBoost	80	1.813554688
	70	-3.995365603
	60	-1.708619003

Based on the comparison, when a bottom threshold below 50% is applied, the resulting second model generally performs poorly. However, when the threshold is lowered to 40%, the new dataset significantly improves the model's accuracy compared to the initial CNN. This indicates that predictions with lower confidence (below 40%) from the CNN model likely contain misclassified data, which are effectively corrected using additional features such as cloud and shadow distances. The experimental results confirm that a lower bottom threshold in retraining the second model leads to better predictions, while a higher top threshold (e.g., above 80%) used to identify temporary cloud and shadow objects further enhances performance. Among the models tested, KNN consistently outperforms RF and XGBoost in terms of accuracy. This suggests that the data features generated especially the spatial distances—are more suitable for algorithms like KNN that rely on proximity and similarity. XGBoost's slightly better performance than RF also aligns with findings from previous studies [28], which emphasized the importance of incorporating advanced feature relationships to handle large, complex datasets. In this research context, the superior performance of KNN reinforces that the second dataset, enriched with cloud and shadow distances, contains structured patterns that are effectively captured through neighborhood-based methods. As a result, KNN will be used in an ensemble configuration to contribute to the final prediction output, improving upon the original CNN model.

4.6. Hybrid model analysis

Based on this research with the given that showed with the experimental top and bottom threshold, the hybrid model that has been create successfully can enhance the accuracy of the data. Another experiment on this research implemented to a single scene and the train and test data are being split into proportion 80 : 20, this experiment can be treat as if the data preprocessing is really ideal scenario. The same model creation then created for the single scene from the one of SPARCS data that being used by [2]. The one dataset that being used is LC81830642014203LGN00, the model creation through the same process with the first model is CNN meanwhile the second model used is KNN.

Table 11. New Hybrid Model Prediction for the single dataset LC81830642014203LGN00

Model	Cloud Accuracy	Shadow Accuracy	Other Accuracy
CNN	89.31%	96.97%	76.45%
New Hybrid Model	95.13%	96.08%	81.25%

With the removal of atmospheric distortion, the CNN model's accuracy improved significantly. The CNN architecture used in this case was simpler than CNN Unet, as the feature characteristics of the Landsat-8 SPARCS image LC81830642014203LGN00 differed from those in previous datasets. Additionally, the newly proposed hybrid model demonstrated the ability to improve accuracy, particularly in dealing with imbalanced datasets, by incorporating cascading methods that add features like cloud and shadow distances. This approach aligns with findings from prior research such as Sun et

al., (2024) who highlighted the advantages of using two-step classification strategies and additional spatial features to enhance model performance. Although the hybrid model in this study does not adopt the same complex architecture or preprocessing as previous studies, it achieves competitive results through strategic use of thresholds and distance-based features. Therefore, the use of top and bottom thresholds as part of the cascading method proves to be a key innovation in model design, and the proposed hybrid framework shows potential for application in future studies that aim to generate improved training datasets and enhance prediction performance from initial CNN outputs.

5. Conclusion

The results of the study demonstrate that the implementation of a cascading-based hybrid deep and machine learning model significantly enhances the accuracy of cloud and shadow segmentation in Landsat-8 imagery. By generating two new features distance to cloud and shadow derived from softmax outputs of a CNN model, the cascading process enables the creation of a refined secondary dataset. This dataset is then used to train a second model (KNN, RF, or XGBoost), with experimental top and bottom thresholds helping to filter high-confidence predictions and uncertain areas needing further classification. Findings reveal that applying a bottom threshold of 40% and a top threshold between 60% to 90% yields notable improvements in accuracy, particularly when using the KNN model, which consistently outperformed RF and XGBoost. The second-stage model effectively captured misclassified areas from the initial CNN output, validating the benefit of cascading features. When the approach was tested on a single SPARCS scene (LC81830642014203LGN00), the hybrid model achieved superior accuracies 95.13% for clouds and 96.08% for shadows compared to the standalone CNN, further confirming the model's robustness. These results underscore the effectiveness of integrating spatial features and threshold-based filtering in hybrid modeling to address challenges in remote sensing segmentation tasks.

Acknowledgment

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

References

- [1] S. A. Boyle, C. M. Kennedy, J. Torres, K. Colman, P. E. Pérez-Estigarribia, and N. U. de la Sancha, "High-resolution satellite imagery is an important yet underutilized resource in conservation biology," *PLoS One*, vol. 9, no. 1, p. e86908, 2014. <https://doi.org/10.1371/journal.pone.0086908>
- [2] M. J. Hughes and R. Kennedy, "High-quality cloud masking of Landsat 8 imagery using convolutional neural networks," *Remote Sens (Basel)*, vol. 11, no. 21, p. 2591, 2019. <https://doi.org/10.3390/rs11212591>
- [3] Z. He, Z. Zhang, M. Guo, L. Wu, and Y. Huang, "Adaptive unsupervised-shadow-detection approach for remote-sensing image based on multichannel features," *Remote Sens (Basel)*, vol. 14, no. 12, p. 2756, 2022. <https://doi.org/10.3390/rs14122756>
- [4] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, pp. 611–629, 2018. <https://doi.org/10.1007/s13244-018-0639-9>
- [5] L. M. Matsunobu, H. T. C. Pedro, and C. F. M. Coimbra, "Cloud detection using convolutional neural networks on remote sensing images," *Solar Energy*, vol. 230, pp. 1020–1032, 2021. <https://doi.org/10.1016/j.solener.2021.10.065>
- [6] E. Ibrahim *et al.*, "Cloud and cloud-shadow detection for applications in mapping small-scale mining in Colombia using sentinel-2 imagery," *Remote Sens (Basel)*, vol. 13, no. 4, p. 736, 2021. <https://doi.org/10.3390/rs13040736>

- [7] R. Li and S. Li, "Multimedia image data analysis based on knn algorithm," *Comput Intell Neurosci*, vol. 2022, no. 1, p. 7963603, 2022. <https://doi.org/10.1155/2022/7963603>
- [8] I. de Zarzà i Cubero, J. de Curtò y DíAz, E. Hernández-Orallo, and C. Calafate, "Cascading and Ensemble Techniques in Deep Learning," *Electronics*, 2023, 12 (15), 1-18., 2023. <https://doi.org/10.3390/electronics12153354>
- [9] B. P. Woolf, "Chapter 7-machine learning," *Building intelligent interactive tutors*, pp. 221–297, 2009.
- [10] H. K. Zhang, D. Luo, and D. P. Roy, "Improved Landsat Operational Land Imager (OLI) Cloud and Shadow Detection with the Learning Attention Network Algorithm (LANA)," *Remote Sens (Basel)*, vol. 16, no. 8, p. 1321, 2024. <https://doi.org/10.3390/rs16081321>
- [11] Z. Wang *et al.*, "Deep learning-based cloud detection for optical remote sensing images: A Survey," *Remote Sens (Basel)*, vol. 16, no. 23, p. 4583, 2024. <https://doi.org/10.3390/rs16234583>
- [12] A. P. Cracknell, "The development of remote sensing in the last 40 years," 2018, *Taylor & Francis*. <https://doi.org/10.1080/01431161.2018.1550919>
- [13] M. A. Ridwan *et al.*, "Applications of landsat-8 data: A Survey," 2018.
- [14] I. W. Gede Astawa Karang and I. G. Hendrawan, "THE USE OF SATELLITE REMOTE SENSING (ALOS SATELLITE DATA)," *International Journal of Remote Sensing and Earth Sciences (IJReSES)*, vol. 4, no. 1, 2010.
- [15] M. M. Taye, "Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions," *Computation*, vol. 11, no. 3, p. 52, 2023. <https://doi.org/10.3390/computation11030052>
- [16] V. Bhosale and Dr. A. Patankar, "Classification of Landsat 8 Images using Neural Networks," *International Journal of Innovative Technology and Exploring Engineering*, vol. 11, no. 8, pp. 23–28, 2022.
- [17] J. Peng, E. C. Jury, P. Dönnies, and C. Ciurtin, "Machine learning techniques for personalised medicine approaches in immune-mediated chronic inflammatory diseases: applications and challenges," *Front Pharmacol*, vol. 12, p. 720694, 2021. <https://doi.org/10.3389/fphar.2021.720694>
- [18] Z. Shao, M. N. Ahmad, and A. Javed, "Comparison of random forest and xgboost classifiers using integrated optical and sar features for mapping urban impervious surface," *Remote Sens (Basel)*, vol. 16, no. 4, p. 665, 2024. <https://doi.org/10.3390/rs16040665>
- [19] G. Zhou and H. Sha, "Building shadow detection on ghost images," *Remote Sens (Basel)*, vol. 12, no. 4, p. 679, 2020. <https://doi.org/10.3390/rs12040679>
- [20] Y. Zi, F. Xie, and Z. Jiang, "A cloud detection method for Landsat 8 images based on PCANet," *Remote Sens (Basel)*, vol. 10, no. 6, p. 877, 2018. <https://doi.org/10.3390/rs10060877>
- [21] D. A. Pitaloka, A. Wulandari, T. Basaruddin, and D. Y. Liliana, "Enhancing CNN with preprocessing stage in automatic emotion recognition," *Procedia Comput Sci*, vol. 116, pp. 523–529, 2017. <https://doi.org/10.1016/j.procs.2017.10.038>
- [22] S. S. Priscila, S. S. Rajest, R. Regin, T. Shynu, and R. Steffi, "Classification of Satellite Photographs Utilizing the K-Nearest Neighbor Algorithm," *Central Asian Journal of Mathematical Theory and Computer Sciences*, vol. 4, no. 6, pp. 53–71, 2023.
- [23] C. Santos and L. Rapp, "Satellite imagery, very high-resolution and processing-intensive image analysis: Potential risks under the GDPR," *Air and Space Law*, vol. 44, no. 3, 2019. <https://doi.org/10.54648/aila2019018>
- [24] A. da P. Pacheco, J. A. da S. Junior, A. M. Ruiz-Armenteros, and R. F. F. Henriques, "Assessment of k-nearest neighbor and random forest classifiers for mapping forest fire areas in central portugal using landsat-8, sentinel-2, and terra imagery," *Remote Sens (Basel)*, vol. 13, no. 7, p. 1345, 2021. <https://doi.org/10.3390/rs13071345>
- [25] Z. Iqbal *et al.*, "Accurate real time localization tracking in a clinical environment using Bluetooth Low Energy and deep learning," *PLoS One*, vol. 13, no. 10, p. e0205392, 2018. <https://doi.org/10.1371/journal.pone.0205392>

- [26] V. Constantinou *et al.*, “Leveraging deep learning for high-resolution optical satellite imagery from low-cost small satellite platforms,” *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 17, pp. 6354–6365, 2024. [10.1109/JSTARS.2024.3365417](https://doi.org/10.1109/JSTARS.2024.3365417)
- [27] M. Fayaz, J. Nam, L. M. Dang, H.-K. Song, and H. Moon, “Land-cover classification using deep learning with high-resolution remote-sensing imagery,” *Applied Sciences*, vol. 14, no. 5, p. 1844, 2024. <https://doi.org/10.3390/app14051844>
- [28] R. P. Dsouza and G. N. K. S. Babu, “Integrating Decision Tree and KNN Hybrid Algorithm approach for Enhancing Agricultural Yield Prediction,” *Computer Integrated Manufacturing Systems*, vol. 30, no. 7, pp. 15–32, 2024.