SPATIAL INVENTORY OF SOLAR PHOTOVOLTAIC (PV) INSTALLATIONS USING REMOTE SENSING AND MACHINE LEARNING: CASE OF PAMPANGA PROVINCE, PHILIPPINES

A. G. Dalagan [1] *, J. A. Principe [2]

[1] National Graduate School of Engineering, University of the Philippines Diliman, Quezon City - acdalagan@up.edu.ph
[2] Dept. of Geodetic Engineering, University of the Philippines Diliman, Quezon City - japrincipe@up.edu.ph

ABSTRACT:

In recent years, the cost of photovoltaic (PV) technologies significantly declined, boosting the expansion of solar PV installations in the country. This rapid development in solar PV utilization necessitates an effective detection method capable of delineating both utility-scale and distributed PV installations to generate a complete inventory of solar PV installations for status monitoring and implementation of appropriate programs of stakeholders and decision-makers. This study aims to detect and delineate solar PV installations in Pampanga, Philippines using different band combinations of Sentinel-2, Sentinel-1, and indices (NDVI, NDWI, and PVSI) through machine learning with the aid of open-source geographic information system and remote sensing software. Moreover, an alternative approach to identifying solar PVs using the combination of pixel-based classification (PBC) and object-based classification (OBC) was introduced. Training and validation data were acquired from the satellite images. The accuracy of each approach in PV detection was then compared using three classifiers: Support Vector Machine (SVM), Random Forest, and Naive Bayes. Results showed that SVM has the best performance for PBC while Random Forest demonstrated the highest accuracy for OBC. A post-processing procedure was also implemented using a set of spectral rules to further refine the results of image classification. The delineation accuracies of the post-processed data over the ground-truth data showed that the methodology is effective in delineating utility-scale installations for both Sentinel-2 and Planetscope. However, the detection of distributed PV systems showed limitations particularly when dealing with small solar PV installations (less than 45 pixels) due to Sentinel-2's coarse spatial resolution.

1. INTRODUCTION

Energy is a crucial component of economic development. As economies grow, so does their dependence on energy consumption. At the same time, the production and utilization of energy somehow negatively impacts the environment. Proper policies and enhanced implementation of sustainable energy are necessary for sustainable economic development (Shengjuan & Jingping, 2011).

The Philippines is vulnerable to the effects of climate change (National Integrated Climate Change Database Information, 2022) due to its geography, high population growth, man-made activities, and exposure to natural hazards (Asian Development Bank, 2018). Since 2008, the country has seen a doubling of greenhouse gas emissions with the energy sector contributing the largest under the Business-As-Usual (BAU) scenario (Department of Energy, 2020). To mitigate this, the adoption of the Paris Agreement aims to reduce greenhouse gas emissions by 75% until 2030 (Aguilar, 2021). To achieve this goal, the Philippine government has implemented various policy options such as carbon taxes, energy efficiency guidelines, adjustments in the energy supply mix, and the proliferation of renewable energy systems (Cabalu et al., 2015).

Energy security is also a continuing problem in the country. Despite improvements in access to electricity, the Philippines still faces problems with energy security due to electric power instability and inaccessibility (World Bank, 2021). Another concern in relation to energy security is self-sufficiency, which refers to the ability to meet energy demand using domestic resources (Kanchana & Unesaki, 2015). In 2020, the country had

a self-sufficiency rate of 53%. However, this rate is expected to decline to 44% in 2030 and 39% in 2040 due to the depletion of Malampaya gas reserves (Department of Energy, 2020). As nonrenewable resources like coal and natural gas continue to deplete, it becomes more challenging to meet the increasing energy demand. Consequently, the country relies heavily on imported fuels, with almost half of the total energy supply in 2020 being imported coal and natural gas. This heavy reliance on fossil fuels leads to high electricity prices, making the Philippines have the highest electricity rate in Southeast Asia (Yokota & Kutani, 2018).

With the increasing electricity consumption and demand and the negative impact of nonrenewables on climate change, the Philippine government is committed to transitioning to cleaner energy resources to support socio-economic development (Asian Development Bank, 2018). Renewable energy is beneficial to the environment as it lowers emissions of GHG and other pollutants (Shinn, 2018). Compared to natural gas and coal, renewable energy can reduce emissions a hundredfold. Transitioning to renewable energy sources can bring about a dual benefit of mitigating the environmental harm caused by energy production while simultaneously boosting energy output and enhancing energy security and stability for the country. Shifting to renewable energy not only reduces harm to the environment but also increases energy stability and security (Centeno, 2018).

Solar energy, which can be harnessed from the sun's radiation, is one of the clean energy sources that can be converted into electricity or other forms of energy. It is recognized as one of the cleanest and richest types of renewable energy. There are two

---

* Corresponding author

ways to harness solar energy: photovoltaic (PV) and concentrated solar power (CSP) (SEIA, 2018). The global installed capacity for solar continues to increase every year, growing by an average of 13% from 2020-2030 due to extensive accessibility of resources, lower costs, and good policy support (International Energy Agency, 2020).

There were two types of PV systems depending on the project size: utility-scale solar installations and distributed PV solar systems. Utility-scale solar installations, also known as solar farms or solar parks, are ground-mounted PV systems found mostly in open areas (Daniels & Wagner, 2022), operated by large energy producers to feed the grid supply, and developed to cater large-scale economies (Donnelly-Shores, 2013). These systems have a capacity of at least 1 MW (Asian Development Bank, 2014) and typically cover 2 to 4 hectares of land depending on the solar technology used (Daniels & Wagner, 2022). Distributed generation systems are small-scale electricity production facilities installed close to the end user of power for onsite consumption (SEIA, 2023). They can be installed on rooftops or as ground-mounted and are usually connected to the local grid. They are intended for small, moderate, or localized use of electricity, such as for residential rooftop and commercial buildings. Some distributed systems may also operate independently from the grid, such as off-grid ground-mounted solar projects that power rural areas or isolated small communities (Hernandez et al., 2014).

The Philippines has a rich solar energy potential with an annual average of 5.1 kWh per square meter per day, based on the assessment of DOE and the US National Renewable Energy Laboratory (IRENA, 2017, p. 20). Given its favorable location, solar energy is a viable source of energy for the Philippines (GIZ, 2013). Solar energy has been rapidly increasing in popularity in the Philippines since 2014, specifically in the small-scale, commercial, and industrial sectors, due to the decreasing cost of solar technologies (Department of Energy, 2020). The use of solar PV systems has resulted in a 10.2% increase in power generation output from solar resources from 2019 to 2020. In 2020, the majority of the increase in renewable energy capacity was due to solar energy, wherein two solar projects, one in Concepcion, Tarlac and the other in Cadiz, started operations, producing 100.6 MW and 132.5 MW of solar power, respectively.

By 2040, the target share of renewable energy is 35% of the total installed capacity (Department of Energy, 2020). To properly plan for future energy efficiency and conservation programs, it's important to have an updated inventory of solar power systems, including distributed rooftop PVs (Kausika et al., 2021). Detecting solar PV using satellite or aerial imagery can help determine location, installed capacity, and area, aiding in the planning of distribution lines (Bradbury et al., 2016) and forecasting energy demand. This information can prevent power interruptions and ensure reliable and efficient power systems (de Hoog et al., 2020). Estimating solar PV panel installations is often inefficient due to lengthy field inspections and manual rooftop examinations. Remote sensing imagery and machine learning can enhance accuracy by detecting solar installations across extensive areas while reducing human error.

The boost in solar panel deployment raised the demand for finding methods for the detection of solar panels. Although there are studies that deal with the solar PV monitoring using machine learning algorithms, there are still limited studies on the detection of these installations for large-scale areas. Solar panel detection has been mainly dominated by efforts that are focusing on rooftop installations in smaller areas using high spatial resolution satellite images. Souffer et al., (2021) used object-based image analysis (OBIA) and machine learning (Random Forest) on UAV RGB and thermal infrared (TIR) images for the extraction of PV panels which yielded 99.7% classification accuracy. Meanwhile, Malof et al., (2015) performed an automated solar panel detection on a portion of Fresno City in California using several methods including SVM classification, MSER-based colour segmentation and shape analysis, boosted cascade classifier and aggregate channel features technique. Both studies obtained good results, but the detection is concentrated on smaller regions.

There were also studies that explored delineation of solar installation for large scale areas. Vasku (2019) compared the classification accuracy of solar farms on Sentinel-2 and Sentinel-1 data using machine learning operations for the entire Denmark. On the other hand, Plakman et al. (2022) identified solar parks using Sentinel-2 and Sentinel-1 data using Random Forest (99.97% accuracy) and experimented on subsampled, oversampled binary data and multiple land cover categories. Although both have successfully detected solar PV installations, the classification is only for utility-scale installations and not distributed PVs (rooftop solar) due to the limitation of using a moderate-resolution image.

Accomplishing a comprehensive solar PV installation database requires not only the detection of large-scale solar farms but also the isolated solar rooftops. Investigating the performance of different multi-resolution data can help identify the best spatial resolution that can consider the detection of both solar PV system types. A study by Jiang et al. (2021) utilized three different spatial resolutions (0.1m, 0.3m, and 0.8m) for solar panel detection. However, these images could still be considered as high-resolution images. Moreover, the study only focused on small areas (the maximum area is 1 km^2). This study will address the above-mentioned gaps and will explore the performance of solar panel detection using satellite images with different spatial resolutions (moderate and high resolution) through machine learning techniques.

The study's objective is to detect and delineate solar PV installations in Pampanga, a province in the Philippines, with the application of machine learning (ML) in different band combinations of Sentinel-2 and Sentinel-1, and indices such as Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Photovoltaic Spectral Index (PVSI) by Shimada & Takeuchi (2022). Solar arrays will also be detected using the Planetscope images. In this study, the performance of various ML classification techniques will be evaluated to identify the best model for the detection of both utility-scale and distributed/rooftop solar PV systems. Also, a classification system comparable to the methods of Guan et al., (2017) is also proposed in this study using the combined pixel-based and object-based classification (POBC) approach. POBC was compared to the traditional pixel-based classification (PBC) and object-based classification (OBC) methods of solar PV panel detection.

2. METHODS

2.1 Study Area

Pampanga is located in Central Luzon (Region 3) on the island of Luzon, Philippines. The region is the second largest contributor to the power supply coming from solar energy systems, according to DOE (2020). In 2019, the power generated

from solar energy in the region is estimated to be about 293,000 GWh.

Furthermore, the region is also a hotspot for solar energy potential, with provinces such as Tarlac, Nueva Ecija, and Pampanga receiving the highest shortwave radiation. (Bauzon et al., 2022). The wide presence of solar PVs in the area and its good geographical location for solar energy are the basis for the selection of the study area.

2.2 Data and Software

The study utilized medium to high-resolution satellite images, consisting of the following:

- Medium-resolution (10-m): Sentinel-2 (Sentinel-1, NDVI, NDWI, and PVSI bands will be incorporated as composite bands), and
- High-resolution (3-m): Planetscope.

Sentinel-2 Level 2A images were acquired in Copernicus Open Access Hub covering Pampanga. The images consist of thirteen bands with a dynamic range of 12 bits per pixel. In addition, Sentinel-1 SAR scenes of Interferometric Wide (IW) swath mode and Ground Range Detected (GRD) type were gathered which are all into descending orbit type to ensure that the data regarding the object's positioning and orientation on the ground is retained. All Sentinel data were captured in February 2019. Meanwhile, Planetscope surface reflectance images were acquired freely through Planet's Education and Research Program for dates February 19, 22, and 23, 2019, covering areas in Pampanga and Tarlac with minimal cloud presence.

The Humanitarian OpenStreetMap Team's 2020 building footprint from OCHA Philippines was utilized to assist in the post-processing phase of image classification. Also, the list of solar PV installations (utility-scale PVs only) (Department of Energy, 2022) was used to facilitate the manual delineation of solar systems.

Various software applications such as SNAP, GRASS GIS, and QGIS including plugin tools like the Semi-Automatic Classification Plugin, Orfeo Toolbox, SNAP, and GRASS GIS were utilized to process and classify satellite images. The supervised machine learning classification operations used were SVM, Random Forest, Decision Tree, and Naïve Bayes.

2.3 Methodology

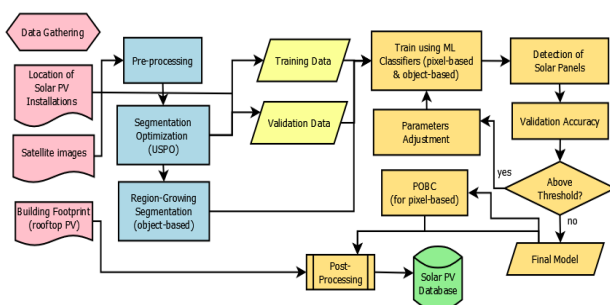The general workflow of the study is illustrated in Fig. 1.

Figure 1. Methodology.

For the preprocessing, the Sentinel-2 and Planetscope images were scaled to get the actual reflectance values. Final input images for classification were obtained by mosaicking, clipping, and applying cloud and shadow masking.

This study compared the effectiveness of integrating different bands for identifying solar PVs. Sentinel-2 was combined with NDVI, NDWI, and PVSI, generating a raster layer for input to the classification process:

$$NDVI = \frac{NIR\ (B8) - Red\ (B4)}{NIR\ (B8) + Red\ (B4)} \tag{1}$$

$$NDWI = \frac{Green\ (B3) - NIR\ (B8)}{Green\ (B3) + NIR\ (B8)} \tag{2}$$

$$PVSI = \frac{2.3*B11 - 1.1*B12 - B8}{2.3*B11 + 1.1*B12 + B8} + 0.5*(B2 - B4 - B8) \tag{3}$$
$$+ signum(1.3 - \frac{B6}{B8}) - 1$$

Sentinel-1 images were also combined with Sentinel-2 bands after being pre-processed and calibrated for radiometric calibration and terrain correction.

Image datasets were split into two categories: training and validation data. The training data is used as a guide for the model to analyze patterns and relationships between the classes in the image, while the validation data helps to fine-tune the model parameters and prevent overfitting/underfitting. Once the model is finalized. Google Earth was used to identify solar PV installations particularly the distributed PVs for creating the training and validation data. The training and validation data classes are composed of solar PV, vegetation, buildings/built-up, bare soil, water, aquafarm, roads, and agricultural lands. 459 polygons were generated for pixel-based classification wherein 20% of the data was used for validation. Object-based classification used the same training data locations as pixel-based classification.

The study implemented two classification types, pixel-based and object-based, as well as an alternative approach of combining both for solar PV detection.

2.3.1 Pixel-based classification

The Orfeo Toolbox plugin in QGIS was used for pixel-based classification. Three classification techniques were used on each image type using training data, Support Vector Machine (SVM), Random Forest (RF), and Naive Bayes (NB). For SVM, the kernel type used was radial basis function while the parameter optimization was activated for the cost and gamma parameter determination. Meanwhile, the selected parameters for the random forest were a *maximum tree depth* of 17 and a *maximum no. of trees* of 1000. Lastly for Naïve Bayes, all parameters were set into the default values.

2.3.2 Object-based classification

To classify an object in an image, the first step is to create segments. The optimal parameters for image segmentation were determined using Unsupervised Segmentation Parameter Optimization (USPO) in GRASS GIS. USPO considers the trade-off between uniformity of objects or intra-segment variance (WV) and spatial autocorrelation (SA) between adjacent objects (Grippa, et al., 2017; Johnson et al., 2015) which is represented as:

$$F = (1 + \alpha^2) * \frac{SA_{norm} * WV_{norm}}{\alpha^2 * SA_{norm} * WV_{norm}} \qquad (4)$$

F-function is used to assess the quality of the segmentation, with values that range from 0 (worst quality) to 1 (best quality), while α indicates weight prioritization to either SA or WV. To lessen the computational complexity of the segmentation process, at least 10% of the total area that generally represents the heterogeneity of the entire image was used as an input for the USPO similar to (Grippa et al., 2017).

Two GRASS GIS modules were used for image segmentation. First, the *i.segment.uspo* module was used to determine the range of threshold to test, threshold interval, and the alpha value for the F-function. The weight prioritization value was selected to produce more over-segmented results, enabling the segmentation of small PV areas.

Second, the *i.segment* module was used for the actual segmentation of the three satellite images using the region-growing segmentation algorithm which requires inputs such as the threshold and min. size parameters. The *threshold* parameter was used to establish the tolerance level for the adjoining segments to be merged based on their closeness in feature space while the *min. size* parameter sets the minimum number of cells in a segment and is selected based on the minimum mapping unit (MMU). The smallest PV array can typically have an area of 9.6 m^2 thus, the min. size parameter used for Sentinel-2 is 1 pixel. In the case of Planetscope, the *min. size* parameter used is 2 pixels considering the large computational power for a lower parameter value. Other USPO input values used were a threshold range of 0.004 to 0.03 with 0.001 step as used by Grippa et al., (2017) and a weight prioritization of 1.25 to place significance on the inter-segment heterogeneity of the image. The final derived optimum segmentation threshold was 0.012 and 0.019 for Sentinel-2 and Planetscope, respectively.

For training the vector classifier model, the parameters selected were similar to PBC.

2.3.3 Accuracy Assessment

To assess the accuracy of each model, several accuracy metrics were performed and compared for the different spatial resolutions of images. Fundamental accuracy metrics such as confusion matrices, overall accuracy and kappa coefficient were obtained for all the classified images.

Other accuracy metrics such as the Jaccard Index or IoU (Intersection over Union), Recall, Precision and F1 Score will also be used to evaluate each machine learning classifier. F1 score (Dice loss) is defined as the harmonic mean of precision and recall. Precision, which is the ratio of the correct positive predictions to the total positive predictions, is also called the user's accuracy for the true positives. Meanwhile, recall, or the fraction of the correct positive predictions from the total true positives is the producer's accuracy of the positives. On the other hand, IoU (Jaccard Index), is the intersection of the ground and classified pixels or objects over the union of the two data groups. It is commonly used as a threshold for identifying true positives and false negatives, with 0.5 as the most commonly preferred threshold (Maxwell et al., 2021)

After obtaining the different accuracy metrics for each type of spatial image, the results together with the visual analysis of each classification will be compared to determine the best-performing model for each spatial resolution for the final post-processing of the solar PV classification in the study area.

2.3.4 Pixel-to-Object-Based Classification

The generated PBC outputs showed many false positives from the misclassification of some pixels in the solar PV class. To mitigate this, the Pixel-to-Object based classification (POBC) process was developed by integrating PBC and OBC to detect solar panels in large-scale regions. The process involves filtering the output from PBC to identify solar PV pixels, converting them to polygons, and treating them as objects for further classification through OBC. This approach eliminates noise from the initial classification and refines the detection process. The results were further refined using the post-processing tasks detailed in the next section.

2.3.5 Post-processing/Refinement

The classified solar pixels were post-processed to remove misclassifications, involving a series of procedures such as reclassification into a binary image and sieving to remove isolated pixels.

A set of spectral rules was then applied similar to the study by Czirjak (2017). Initially, the spectral reflectance range of solar PVs were obtained to remove highly reflective pixels. Only the NIR region was filtered for Sentinel-2 as the solar PV's spectral profile exhibits noticeable separation compared to other classes within this spectral range. The second filter used a threshold determined by 15% of the lowest PVSI value of the sample data. The third filter removes pixels with wide variability in reflectance, considering the consistent reflectance pattern of the solar PV within the visible region. All pixels with standard deviation in the visible region should be lower than the product of the threshold and mean of the visible bands. The threshold described is the adjusted average value of the normalized standard deviations.

Lastly, the building footprint was used to filter out smaller solar pixels found in built-up areas. To avoid elimination of utility-scale solar PVs since it is commonly located around agricultural lands, only areas less than 20,000 m^2 were considered for the filtering.

3. RESULTS

3.1 Classification Results via Visualization

3.1.1 Sentinel-2

Fig. 2 shows the sample detected solar PV installations using the three classifiers via pixel-based classification. For the sample detected utility-scale solar PV, it can be observed that almost similar results were obtained, and the classifiers generally obtained the outline of the solar farm. SVM and Random Forest showed excellent results in the classification. Meanwhile, Naive Bayes failed to identify portions of the solar farm, especially the narrower regions.

For the performance of Sentinel-2 combinations, the stacked Sentinel-2 with indices appears to have the best classification output as compared to the Sentinel-2 and Sentinel 1 which misclassified some pixels in the proximity of the solar farm.
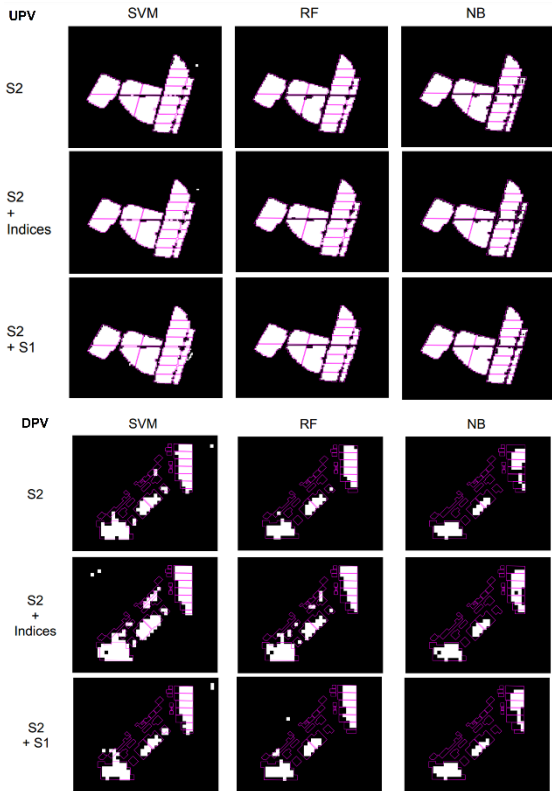
Figure 2. Detected solar PV installations from the PBC of different Sentinel-2 image combinations.

Figure 3. Detected solar PV installations from the OBC of different Sentinel-2 image combinations.

In the case of distributed solar PVs (DPVs), the classification of the panels depends on the size of the installed PV array on the rooftop. Solar installations with larger areas were able to be delineated by the classifiers specifically, SVM and Random Forest. However, those rooftop solar with smaller areas and a larger gap with nearby rooftop solar were not detected by the algorithm. Visibly, it is Sentinel-2 with the indices, namely NDVI, NDWI and PVSI, that were able to best capture the actual solar PV footprint in the area.

The sample detected utility-scale and distributed solar PV installations through OBC is shown in Fig. 3. As compared to PBC, there was more noise generated on the solar classification outputs for OBC, especially for the combination of Sentinel-2 and Sentinel-1. Moreover, Random Forest produced the best detection results among all the classifiers since it was able to avoid the misrecognition of the gaps between each group of solar arrays. Similar to PBC, UPV can be accurately detected using the Sentinel-2 with the indices composite.
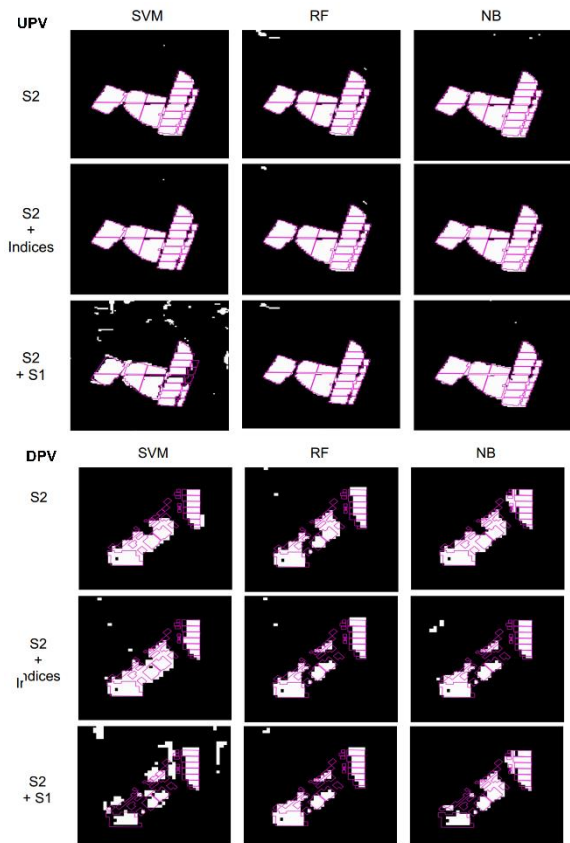
For the OBC of the DPVs, it is apparent that the solar rooftop areas in the lower left portion were merged for the SVM algorithm. There are also misclassified objects around the area in contrast with PBC. Same with the UPVs, Random Forest yielded the best visual output from all the detections.

3.1.2 Planetscope

For the Planetscope image, Fig. 4 show the detected solar PVs for both PBC and OBC, respectively. Similar to Sentinel-2, SVM demonstrated the closest visual resemblance to the actual solar PV footprint for PBC and Random Forest for OBC. It is observed that there is a greater prevalence of false positives around the distributed PVs in Planetscope data compared to the Sentinel-2 PV detections, with the worst occurrence exhibited in the Naive Bayes classification. This problem is common with high spatial resolution data due to its high spectral variation, making it susceptible to spectral confusion among relatively similar pixels (Moran, 2010).
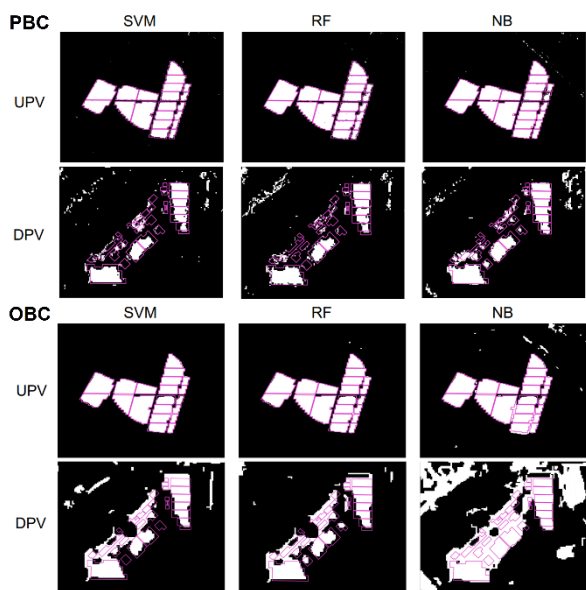
Figure 4. Detected solar PVs of Planetscope image.

3.2 Accuracy Assessment using Different Accuracy Metrics

The training accuracies for Sentinel-2, Sentinel-2 composites with Sentinel-1 and the indices, and Planetscope using the two classification approaches are shown in Table 1. Accuracy metrics such as precision, recall, and F1-score for the solar PV class were introduced for all the classifiers used. The accuracy of the whole classification was also measured through the Kappa metric and overall accuracy. Evidently, all Sentinel-2 image combinations showed a good performance for classifying solar PV panels.

Comparing the accuracy values for the classification of three Sentinel-2 variants, Sentinel-2 with Sentinel-1 bands demonstrated the highest Kappa and overall accuracy values. However, Sentinel-2 with the indices provided the best F1 score for solar PV classification. SVM yielded the highest values for all the accuracy metrics used.

PBC				
Accuracy Metrics	Sentinel-2	Sentinel-2 + Indices	Sentinel-2 + Sentinel-1	Planet-scope
Precision (solar PV)	SVM: 99% RF: 99% NB: 99%	SVM: 99% RF: 98% NB: 99%	SVM: 99% RF: 99% NB: 99%	SVM: 97% RF: 94% NB: 85%
Recall (solar PV)	SVM: 99% RF: 97% NB: 95%	SVM: 99% RF: 98% NB: 97%	SVM: 98% RF: 98% NB: 96%	SVM: 96% RF: 94% NB: 90%
F1 Score (solar PV)	SVM: 99% RF: 98% NB: 97%	SVM: 99% RF: 98% NB: 99%	SVM: 99% RF: 99% NB: 98%	SVM: 96% RF: 94% NB: 88%
Kappa	SVM: 0.96 RF: 0.94 NB: 0.85	SVM: 0.97 RF: 0.94 NB: 0.87	SVM: 0.99 RF: 0.97 NB: 0.91	SVM: 0.86 RF: 0.84 NB: 0.71
Overall Accuracy	SVM: 96% RF: 95% NB: 87%	SVM: 97% RF: 95% NB: 89%	SVM: 99% RF: 97% NB: 91%	SVM: 87% RF: 86% NB: 75%

OBC				
Accuracy Metrics	Sentinel-2	Sentinel-2 + Indices	Sentinel-2 + Sentinel-1	Planet-scope
Precision (solar PV)	SVM: 99% RF: 99% NB: 92%	SVM: 99% RF: 99% NB: 95%	SVM: 97% RF: 99% NB: 93%	SVM: 91% RF: 99% NB: 30%
Recall (solar PV)	SVM: 97% RF: 97% NB: 99%	SVM: 99% RF: 99% NB: 99%	SVM: 92% RF: 99% NB: 99%	SVM: 84% RF: 86% NB: 95%
F1 Score (solar PV)	SVM: 99% RF: 99% NB: 96%	SVM: 99% RF: 99% NB: 98%	SVM: 95% RF: 99% NB: 96%	SVM: 87% RF: 86% NB: 45%
Kappa	SVM: 0.94 RF: 0.99 NB: 0.84	SVM: 0.95 RF: 0.99 NB: 0.84	SVM: 0.93 RF: 0.99 NB: 0.89	SVM: 0.84 RF: 0.94 NB: 45%
Overall Accuracy	SVM: 95% RF: 99% NB: 86%	SVM: 96% RF: 99% NB: 86%	SVM: 94% RF: 99% NB: 91%	SVM: 88% RF: 96% NB: 68%

Table 1. Accuracy values of the different Sentinel-2 image combinations and Planetscope.

For the OBC, Sentinel-2 with Sentinel-1 produced the lowest accuracy values among the three Sentinel-2 band combinations in contrast to the PBC results. Meanwhile, Sentinel-2 with the indices consistently showed the best performance among the three image inputs. In terms of accuracy concerning the ML classifier used, Random Forest generated the best classification output in relation to its accuracy values.

The classification of the solar PV class in the entire Pampanga is observed to have worse presence of misclassified pixels in Sentinel-2 with Sentinel-1 bands as compared to Sentinel-2 and indices. Despite the higher accuracy values obtained for PBC, the classification results for the latter are still considered as the better output as it produced fewer misclassifications than the former.

For the finer spatial image, SVM consistently produced the best accuracy among the classifiers for PBC while Random Forest for the case of OBC. All the calculated accuracy metric values of Planetscope also displayed lower values than Sentinel-2, which was primarily due to the numerous misclassifications caused by the spectral complexity of the image. In spite of the lower accuracy values, the detection of the distributed PVs in the Planetscope image was found to be remarkably better than Sentinel-2.

3.3 Application of POBC and Post-Processing Techniques

Considering the preceding outcomes of the different solar PV classifications, SVM and Random Forest were identified as the best models for PBC and OBC respectively, with Sentinel-2 and the indices as the image input.

Fig. 5 shows the detected solar PVs using the final model for both PBC and OBC for Sentinel-2 with the indices as well as the post-processed classifications. Red areas are the actual solar PV delineations while the white pixels/objects show the misclassified solar PV regions. For the raw classification, both approaches exhibited substantial noises which are typically clustered in built-up areas and roads.

After the application of the post-processing procedures, Fig. 5c and Fig. 5d shows the generated output for both PBC and OBC.

A significant number of false positives were eliminated after implementing the post-processing steps, but some incorrect detections remain in built-up areas.
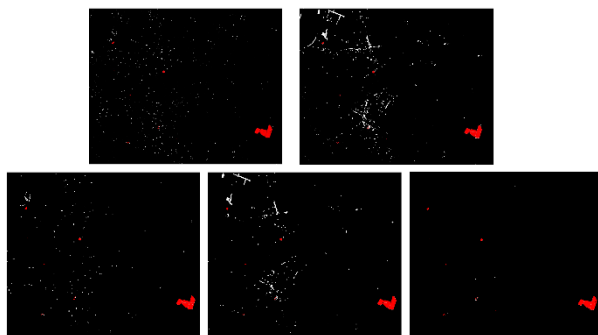

Figure 5. Raw classification of solar PVs using *(from left to right)* (a) PBC (b) OBC; After post-processing (c) PBC (d) OBC and (e) POBC.

The implementation of OBC after PBC, along with the subsequent post-processing has effectively eradicated almost all the false positives as shown in Fig. 5e.

3.4 Delineation Accuracy (Jaccard Index)

To compare the delineation accuracy of the post-processed images, the Jaccard Index (IoU) was computed from the results of both the best-performing model for Planetscope and Sentinel-2. From Table 2, POBC derived the highest IoU value for the combination of Sentinel-2 with the Indices. Generally, the Sentinel-2 with the indices exhibited excellent results for the utility-scale PVs. Meanwhile, the obtained low Iou values for the DPVs indicate difficulty in accurately delineating small PVs in the images due to low spatial resolution.

UPV	% Overlap	IoU	DPV	% Overlap	IoU
PBC	S2: 96% P: 90%	S2: 0.90 P:0.88	PBC	S2: 40% P: 58%	S2: 0.26 P: 0.33
OBC	S2: 91% P: 94%	S2: 0.83 P: 0.90	OBC	S2: 44% P: 44%	S2: 0.17 P: 0.27
POBC	S2: 96% P: 88%	S2: 0.90 P: 0.87	POBC	S2: 51% P: 52%	S2: 0.31 P: 0.39

Table 2. IoU values of the Sentinel-2 with Indices and Planetscope.

Different results were observed for the Planetscope images. While there are minimal differences among the IoU values for UPV, OBC demonstrates the highest similarity value compared to the other approach. For DPV, POBC still has the highest degree of similarity as compared with the ground truth data. The results also imply that Planetscope demonstrated better delineation of the distributed PVs compared to Sentinel-2.

3.5 Detected Solar PVs

Table 3 tabulates all the solar PVs detected in Pampanga obtained using the POBC approach. Fourteen (14) solar PVs were identified wherein 2 are utility-scale PVs while the rest are distributed PVs. Only 17% of the distributed PVs showed good delineation but the acquired detection rate is 93%.

id	Type	Area (m²)	Detected Area in m² (S2)	Detected Area in m² (P)	IoU (S2)	IoU (P)
1	distributed PV	210.71	194.08	40.55	0.27	0.18
2		212.78	51.43	22.27	0.20	0.07
3		249.46	131.75	37.36	0.32	0.13
4		307.59	128.72	260.99	0.34	0.34
5		355.58	283.39	165.52	0.37	0.33
6		526.92	283.49	70.89	0.34	0.12
7		549.20	126.27	246.84	0.10	0.31
8		707.92	395.63	618.68	0.22	0.70
9		954.86	n/a	359.68	n/a	0.38
10		2,464.06	829.59	1,516.48	0.25	0.50
11		3,085.86	900.10	1,551.25	0.26	0.40
12		4,470.78	2,852.06	2,910.25	0.50	0.53
13		18,821.77	12,699.90	8,543.87	0.59	0.44
14	utility scale PV	180,732.52	175,507.79	160,996.47	0.92	0.89
15		193,139.95	185,096.62	169,498.10	0.88	0.86

Table 3. Detected solar PVs for Sentinel-2 and the indices and Planetscope.

Fifteen (15) solar PVs were detected for the Planetscope image (POBC), with an additional distributed PV identified during the classification process. 20% of the distributed PVs have obtained an acceptable delineation accuracy while the detection rate achieved is 85%.

4. DISCUSSION AND CONCLUSION

Results have shown that the model trained using SVM achieved superior performance for the detecting solar installations using PBC, while Random Forest performed better for the case of OBC.

Among the Sentinel-2 composites, Sentinel-2 with the indices has provided the best classification output. In comparison with the pixel-based classification conducted by Vasku (2019), similar pattern was observed, with both SVM and Random Forest generating good accuracy results in delineating utility-scale solar PVs and the incorporation of Sentinel-1 with Sentinel-2 performing better than using Sentinel-2 alone.

Plakman et al., (2022) which employed Random Forest via OBC (utilized Simple Non-Iterative clustering segmentation) acquired different outcomes from this study, with Sentinel-2 and Sentinel-1 composite displaying higher accuracy than the combinations with indices (Sentinel-2 + NDVI and Sentinel-2 + NDWI). In this study, the use of Sentinel-2 with indices provided the highest accuracy values for OBC using the same classifier. In contrast with the methods by Plakman et al., (2022), this study proposed a distinctive approach of combining an additional solar PV spectral index with Sentinel-2 and the other indices. This factor, along with the differences in the segmentation process and the type of classes, likely contributed to the observed differences in results.

Interestingly, an existing study by Guan et al., (2017) explored the use of successive classification stages through the combination of pixel-based classification using SVM, multi-feature classification, and object-based classification in rooftop solar detection using high resolution imagery. Multi-feature classification categorized non-solar features based on the identified correlation between the blue band and red band of a rooftop solar and regular building. Results revealed that the object-based classification process eliminated some small PV arrays from the final detection map. Similar observation was inferred in this study, however, the said drawbacks in using POBC had predominantly affected utility-solar PV installations, reflecting in the minor difference in the delineation accuracies between PBC and POBC. The decrease in delineation accuracy can be attributed to the elimination of pixels in the object-based classification process of POBC around the boundary of the solar PVs, as these pixels appears to be "mixed pixels", or pixels that represents a mixture of land cover classes. Notably, the proposed POBC demonstrated the best delineation accuracy for the detection of distributed solar PVs.

To identify the number of PVs with good delineation accuracy, a threshold of 0.5 was used to categorize the IoU values. Evidently, all UPVs yielded high delineation accuracies. Meanwhile, the DPVs exhibited varying performance in the delineation depending on the actual area of the solar PV. Only two DPVs displayed good IoU values for Sentinel-2 while, three DPVs in Planetscope resulted in satisfactory delineations. In terms of area, Sentinel-2 can accurately delineate areas larger than 4,500 m^2 and 710 m^2 for the case of Planetscope.

As previously stated, the percentage of the total detected area of Sentinel-2 with indices over the total actual area (detection rate) covered by solar PVs exceeds that of Planetscope. This difference could be attributed to Sentinel-2's higher spectral resolution, allowing for better discernment among confusing pixels. Additionally, it could be hypothesized that the generalization of the pixels in Sentinel-2 played a huge factor in the larger area detected. This suggests that despite the higher detection rate obtained, Sentinel-2 might have included pixels beyond the actual PV boundary which explains the lower similarity index values generated as compared to Planetscope.

At present, the available solar PV data in the country remains incomplete specifically in the case of distributed PV installations. Having a comprehensive PV installation is imperative for effective future planning in solar development. Existing methods for solar PV detection either focused on identification of rooftop solar using high-spatial resolution in limited area or delineating utility-scale PVs in large-scale area using moderate image resolution. This study introduced a semi-automated method of detecting both PV types, using three different approaches applied with subsequent post-processing techniques.

Overall, the application of machine learning in the detection of solar PVs using Sentinel-2 and Planetscope is proven to be highly effective, particularly for solar farms and distributed PVs covering large areas. Planetscope exhibited a better performance in locating and delineating distributed PVs as compared to Sentinel-2 but demonstrated a slightly weaker performance in delineating utility-scale PVs. The use of classifiers such as SVM and Random Forest has also been proven to worked well and may be employed on further classification of solar PVs on a larger scale. For future studies, a more advanced post-processing method may be developed to further removal of misclassified pixels. Additionally, exploring the use of other indices (e.g. urban) as an additional band may also be considered.

REFERENCES

Aguilar, K. (2021). PH vows to cut gas emissions by 75 percent. Inquirer News. https://newsinfo.inquirer.net/1420004/ph-vows-to-cut-gas-emissions-by-75-percent

Asian Development Bank (Ed.). (2014). *Handbook for rooftop solar development in Asia*. Asian Development Bank.

Asian Development Bank. (2018). Philippines: Energy Sector Assessment, Strategy, and Road Map (0 ed.). Asian Development Bank. https://doi.org/10.22617/TCS189616

Bauzon, M. D. A., Sotto, M., Santos, J. A., & Principe, J. (2022). Spatial Analysis on Solar PV Potential Hotspots using Himawari-8 SWR data and Locations of Existing Solar PV Projects: Case of the Philippines. Proceedings of the 30th IIS Forum, 29–36.

Bradbury, K., Saboo, R., L. Johnson, T., Malof, J. M., Devarajan, A., Zhang, W., M. Collins, L., & G. Newell, R. (2016). Distributed solar photovoltaic array location and extent dataset for remote sensing object identification. Scientific Data, 3(1), Article 1. https://doi.org/10.1038/sdata.2016.106

Cabalu, H., Koshy, P., Corong, E., Rodriguez, U.-P. E., & Endriga, B. A. (2015). Modelling the impact of energy policies on the Philippine economy: Carbon tax, energy efficiency, and changes in the energy mix. Economic Analysis and Policy, 48, 222–237. https://doi.org/10.1016/j.eap.2015.11.014

Centeno, R. (2018). Solar Energy in the Philippines. Stanford. http://large.stanford.edu/courses/2018/ph240/centeno1/

Czirjak, D. W. (2017). Detecting photovoltaic solar panels using hyperspectral imagery and estimating solar power production. Journal of Applied Remote Sensing, 11(2), 026007. https://doi.org/10.1117/1.JRS.11.026007

Daniels, T. L., & Wagner, H. (2022). *Regulating Utility-Scale Solar Projects on Agricultural Land*. https://kleinmanenergy.upenn.edu/wp-content/uploads/2022/08/KCEP-Regulating-Utility-Scale-Solar-Projects.pdf

de Hoog, J., Maetschke, S., Ilfrich, P., & Kolluri, R. R. (2020). Using Satellite and Aerial Imagery for Identification of Solar PV: State of the Art and Research Opportunities. Proceedings of the Eleventh ACM International Conference on Future Energy Systems, 308–313. https://doi.org/10.1145/3396851.3397681

Department of Energy. (2020). Philippine Energy Plan (PEP). https://www.doe.gov.ph/pep

Department of Energy. (2022). Operating Solar Projects Registered with DOE as of March 20, 2022.

Donnelly-Shores, P. (2013, July 30). *What Does 'Utility-Scale Solar' Really Mean?* https://www.greentechmedia.com/articles/read/what-does-utility-scale-solar-really-mean

Energy Saving Trust. (2015). Solar Energy Calculator Sizing Guide. https://pvfitcalculator.energysavingtrust.org.uk/Documents/150224_SolarEnergy_Calculator_Sizing_Guide_v1.pdf

GIZ. (2013). Facts and Figures on Solar Energy in the Philippines Project Development Programme (PDP) Southeast-Asia. Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH. https://www.doe.gov.ph/sites/default/files/pdf/netmeter/policy-brief-its-more-sun-in-the-philippines-V3.pdf?withshield=1

Grippa, T., Georganos, S., Lennert, M., Vanhuysse, S., & Wolff, E. (2017). A local segmentation parameter optimization approach for mapping heterogeneous urban environments using VHR imagery. Remote Sensing Technologies and Applications in Urban Environments II, 10431, 79–97. https://doi.org/10.1117/12.2278422

Guan, X., Qi, W., Zhang, M., Chen, T., He, J., Wen, Q., & Wang, Z. (2017). Post-Processing-Based Solar Photovoltaic Detection in High Resolution Aerial Imagery. 38th Asian Conference on Remote Sensing. ACRS 2017, India. https://a-a-r-s.org/proceeding/ACRS2017/Missing_pro2/Missing_2_180321/1074.pdf

Hernandez, R. R., Hoffacker, M. K., & Field, C. B. (2014). Land-Use Efficiency of Big Solar. *Environmental Science & Technology*, *48*(2), 1315–1323. https://doi.org/10.1021/es4043726

International Energy Agency. (2020). World Energy Outlook 2020. IEA Publications.

IRENA. (2017). Renewables Readiness Assessment: The Philippines. IRENA.

Johnson, B. A., Bragais, M., Endo, I., Magcale-Macandog, D. B., & Macandog, P. B. M. (2015). Image Segmentation Parameter Optimization Considering Within- and Between-Segment Heterogeneity at Multiple Scale Levels: Test Case for Mapping Residential Areas Using Landsat Imagery. ISPRS International Journal of Geo-Information, 4(4), Article 4. https://doi.org/10.3390/ijgi4042292

Kanchana, K., & Unesaki, H. (2015). Assessing Energy Security Using Indicator-Based Analysis: The Case of ASEAN Member Countries. Social Sciences, 4(4), 1269–1315. https://doi.org/10.3390/socsci4041269

Kausika, B. B., Nijmeijer, D., Reimerink, I., Brouwer, P., & Liem, V. (2021). GeoAI for detection of solar photovoltaic installations in the Netherlands. Energy and AI, 6, 100111. https://doi.org/10.1016/j.egyai.2021.100111

Maxwell, A. E., Warner, T. A., & Guillén, L. A. (2021). Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review. Remote Sensing, 13(13), Article 13. https://doi.org/10.3390/rs13132450

Moran, E. Federico. (2010). Land Cover Classification in a Complex Urban-Rural Landscape with Quickbird Imagery. Photogrammetric Engineering and Remote Sensing, 76(10), 1159–1168.

National Integrated Climate Change Database Information. (2022). Climate Change Impacts. NICCDIES. https://niccdies.climate.gov.ph/climate-change-impacts

SEIA. (2018). Solar Energy Technologies. SEIA. https://www.seia.org/sites/default/files/inline-files/SEIA-Solar-Energy-Technologies-Factsheet-2018-April.pdf

SEIA. (2023). *Rooftop Solar*. https://www.seia.org/initiatives/rooftop-solar

Shengjuan, H., & Jingping, Z. (2011). Research on the Dynamic Relationship of the Energy-Economy-Environment (3E) System-Based on an Empirical Analysis of China. Energy Procedia, 5, 2397–2404. https://doi.org/10.1016/j.egypro.2011.03.412

Shimada, S., & Takeuchi, W. (2022). A new spectral index to characterize solar photovoltaic panels for Sentinel-2 data. 43rd Asian Conference on Remote Sensing.

Shinn, L. (2018). Renewable Energy: The Clean Facts. NRDC. https://www.nrdc.org/stories/renewable-energy-clean-facts

World Bank. (2021). Access to electricity (% of population)—Philippines. https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS?locations=PH

Yokota, E., & Kutani, I. (2018). Comparative Analysis of Power Prices in the Philippines and Selected ASEAN Countries (33-64 Chapter 3: Comparing Electricity Cost Structures. Economic Research Institute for ASEAN and East AsiaEconomic Research Institute for ASEAN and East Asia. https://www.eria.org/uploads/media/ERIA_RPR_2017_12.pdf