



Contributing Paper

# Illegal Dumpsite Detection Guatemala

Technical Paper



## **United Nations Development Programme**

The United Nations Development Programme (UNDP) is the leading United Nations organization fighting to end the injustice of poverty, inequality, and climate change. Working with our broad network of experts and partners in 170 countries, we help nations to build integrated, lasting solutions for people and planet.

Learn more at [undp.org](https://undp.org) or follow at [@UNDP](https://twitter.com/UNDP).

## **UNDP Sustainable Finance Hub**

The UNDP Sustainable Finance Hub (SFH) brings together UNDP's financial expertise to harness public and private capital for the Sustainable Development Goals (SDGs) – supporting governments, investors and businesses in reaching climate, social impact and sustainability targets. Its work drives systemic change towards a sustainable financial architecture that benefits people and the planet.

Find out more about its integrated services that ensure all finance is sustainable at [sdgfinance.undp.org](https://sdgfinance.undp.org) or follow [@UNDP\\_SDGFinance](https://twitter.com/UNDP_SDGFinance).

## **UNDP Accelerator Labs**

The United Nations Development Programme (UNDP) Accelerator Labs is the world's largest and fastest learning network on wicked sustainable development challenges. Co-built as a joint venture with the Federal Ministry for Economic Cooperation and Development of Germany and the Qatar Fund for Development, along with Partners at Core for UNDP, the Italian Ministry of Environment and Energy Security as action partner, and the Japan Cabinet, the Network covers 113 countries, and taps into local innovations to create actionable intelligence and reimagine sustainable development for the 21st century.

The Network has been awarded Apolitical's Public Service Team of the Year in 2019, covered in the MIT Sloan Review (2020), analyzed as a Harvard Business School Case Study (2022) and depicted in "for Tomorrow," a documentary on grassroots innovation, available on Amazon Prime.

Learn more at [acceleratorlabs.undp.org](https://acceleratorlabs.undp.org) or follow us at [@UNDPAccLabs](https://twitter.com/UNDPAccLabs)

# Table of Contents

<b>Glossary of Terms .....</b>	<b>5</b>
<b>1 Overview .....</b>	<b>7</b>
<b>2 Background .....</b>	<b>8</b>
2.1 Related work .....	8
<b>3 Key objectives .....</b>	<b>10</b>
3.1 Research questions .....	10
3.2 Key steps .....	10
3.3 Deliverables .....	10
<b>4 Data .....</b>	<b>11</b>
4.1 AerialWaste dataset .....	11
4.2 Ground truth data (Guatemala) .....	12
4.3 Image dataset (Guatemala) .....	13
<b>5 Methodology .....</b>	<b>14</b>
5.1 Model architecture .....	14
5.2 Transfer learning on Guatemala dataset .....	15
5.3 Dataset sampling .....	15
5.4 Validation metrics: Precision, recall and F1 score .....	15
<b>6 Results .....</b>	<b>17</b>
6.1 Final model results .....	19
<b>7 Discussion.....</b>	<b>20</b>
<b>8 Future work .....</b>	<b>21</b>
<b>9 Conclusion.....</b>	<b>23</b>
Appendix: Locations marked as positive in 2022 for illegal dumpsites but lacking visible evidence in satellite imagery.....	24
<b>Bibliography .....</b>	<b>25</b>

# Acknowledgement

The team would like to thank the partner organizations for their engagement and feedback. The UNDP Istanbul Centre for Private Sector in Development (ICPSD) would like to thank the UNDP Guatemala Accelerator Lab for their kind support and contributions. We would also like to express our gratitude to Carlos Estuardo Mazariegos Orellana at the UNDP Guatemala Accelerator Lab for the guidance and insight he offered throughout the research process. Furthermore, we would like to thank Politecnico di Milano (Polimi) for their essential groundwork and technical assistance. We would also like to thank all peer reviewers and editors, designers and facilitators.

## Authors:

Cedric Baron (SDG AI Lab, ICPSD)  
Martin Szigeti (SDG AI Lab, ICPSD)  
Gokhan Dikmener (SDG AI Lab, ICPSD)

## Peer reviewers:

Ceyda Alpay (Head of Experimentation, UNDP Türkiye Accelerator Lab)  
Cem Bayrak (Social Inclusion and Governance Portfolio Manager, UNDP Türkiye)  
Jeremy Boy (Lead Data Scientist, UNDP Accelerator Labs Network)  
Piero Fraternali (Professor of Computer Science, Polimi)

Their thoughtful comments and suggestions have significantly contributed to the rigour and relevance of the paper.

# Glossary of Terms

**AerialWaste dataset:** An open-source dataset consisting of high-resolution aerial images of waste dump locations, primarily from the Lombardy region in Italy.

**AGEA orthophotos:** High-resolution images captured from an airborne campaign by the Italian Agriculture Development Agency, used in the AerialWaste dataset.

**Artificial intelligence:** The simulation of human intelligence processes by machines, especially computer systems, to perform tasks such as visual perception, speech recognition, decision-making and language translation.

**Computer vision:** The process by which artificial intelligence systems analyse and interpret visual data such as images and videos.

**Convolutional neural network (CNN):** A class of deep neural networks commonly used to analyse visual imagery.

**Deep learning:** A subset of machine learning that involves neural networks with many layers, capable of learning from large amounts of data.

**Feature pyramid network (FPN):** A neural network architecture that improves performance in image recognition tasks by utilizing features at multiple levels.

**Geographical information systems (GIS):** Systems designed to capture, store, analyse and display geographic data, enabling users to visualize, interpret and understand spatial relationships in various contexts.

**Google Maps Static API:** A service that allows developers to embed Google Maps images into web pages without requiring JavaScript.

**Ground truth data:** Data collected on-site and used as a reference to validate the results of predictive models.

**Image classification:** The process of categorizing and labelling groups of pixels or vectors within an image based on specific rules.

**ImageNet:** A large visual database used for training image recognition software.

**Machine learning:** A subset of artificial intelligence that enables systems to learn and improve from experience without being explicitly programmed. It focuses on the development of algorithms that allow computers to learn from and make predictions or decisions based on data.

**Precision:** A metric that measures the accuracy of positive predictions, defined as the ratio of true positives to the sum of true positives and false positives.

**Recall:** A metric that measures the ability to identify all relevant instances, defined as the ratio of true positives to the sum of true positives and false negatives.

**ResNet50:** A CNN architecture with 50 layers, known for its use of residual connections to improve training and performance.

**Single shot detector:** An object detection model that is able to detect objects in images in a single forward pass of the network.

**Support vector machine:** A supervised learning model used for classification and regression analysis.

**Transfer learning:** A machine learning technique where a pre-trained model is adapted to perform a different but related task.

**Validation metrics:** Measures used to evaluate the performance of a model, including precision, recall and F1 score.

**WorldView-3 satellite:** A high-resolution commercial earth observation satellite, providing images used in remote sensing applications.

**You Only Look Once (YOLO):** A real-time object detection system that detects objects in images in a single evaluation.



Photo: Envato / Mehaniq4<sup>1</sup>

# 1 Overview

This project aims to identify illegal waste dumping sites in Guatemala through state-of-the-art geographical information systems (GIS), computer vision and machine learning methods. Co led by the UNDP Guatemala Accelerator Lab<sup>1</sup> and UNDP Istanbul Centre for Private Sector in Development (ICPSD) SDG AI Lab<sup>2</sup>, and in collaboration with Politecnico di Milano (Polimi)<sup>3</sup>, the project uses satellite imagery analysis to facilitate efficient and cost-effective monitoring of dumpsites. This joint effort aligns with the strategic goals of the Guatemala National Development Plan and Country Priorities.

---

<sup>1</sup> UNDP Accelerator Labs, <https://undp.org/acceleratorlabs>

<sup>2</sup> SDG AI Lab, <https://sdgailab.org>.

<sup>3</sup> Politecnico di Milano (Polimi), <https://polimi.it>.

## 2 Background

Guatemala is actively addressing the environmental challenges posed by the management of solid waste at the municipal level, as outlined in its National Development Plan and Country Priorities. The Ministry of Environment and Natural Resources and UNDP Guatemala are making efforts to establish baselines and indicators to understand and measure the impact of harmful waste in areas of biological importance, particularly in critical fluvial and marine-coastal zones like the Motagua River. This river, the largest in Guatemala, is of significant concern as it transports a vast amount of human-made waste to the Atlantic Ocean, leading to a geopolitical crisis with Honduras.<sup>4</sup>

To combat this, UNDP Guatemala experimented with digital technologies to enhance the design and implementation of solid waste management strategies. The UNDP Accelerator Lab is at the forefront of this initiative, exploring innovative computer vision tools applied on diverse data sources such as land surface temperature and high-resolution satellite imagery. Advanced GIS and machine learning techniques, particularly convolutional neural networks (CNNs), can help speed up the identification of illegal dumping sites and avoid the cost and time of traditional waste site monitoring methods, which are common constraints for developing countries.

### 2.1 Related work

In the realm of waste detection using remote sensing data, significant progress has been made over the years, with a notable shift from human interpretation of aerial images to the application of advanced computational techniques. Early efforts in the 1970s and 1980s, like those by Garofalo and Wobber<sup>5</sup> and Lyon,<sup>6</sup> focused on using aerial photos to analyse waste distribution and document landfill locations using historical images.

A significant leap occurred in 2017 with the integration of machine learning techniques. Selani<sup>7</sup> achieved 85.16 percent accuracy in classifying high-resolution WorldView-2 multispectral images into categories including waste types using a support vector machine. This study highlighted the varying significance of different spectral bands in identifying various waste categories. Similarly, Angelino et al.<sup>8</sup> employed satellite images in 2018 to identify illegal landfills in southern Italy using photo-interpretation by experts,

---

<sup>4</sup> Jeff Abbott, “Guatemala’s rivers of garbage”, The Progressive, 16 August 2023.

<sup>5</sup> Donald Garofalo and Frank J. Wobber, “Solid waste and remote sensing”, *Photogrammetric Engineering*, vol. 40 (1974), pp. 45–59.

<sup>6</sup> John Grimson Lyon, “Use of maps, aerial photographs, and other remote sensor data for practical evaluations of hazardous waste sites”, *Photogrammetric Engineering and Remote Sensing*, vol. 53 (1987), pp. 515–519.

<sup>7</sup> Lungile Selani, “Mapping illegal dumping using a high resolution remote sensing image: Case study: Soweto Township in South Africa”, PhD dissertation, University of the Witwatersrand, 2017.

<sup>8</sup> Cesario Vincenzo Angelino et al., “A case study on the detection of illegal dumps with GIS and remote sensing images”, in *Earth Resources and Environmental Remote Sensing/GIS Applications IX*, vol. 10790 (SPIE Remote Sensing, 2018).

complemented with a multi-feature detection algorithm. They also performed a multi-temporal analysis using satellite images, which allowed for tracking the evolution of landfills.

More recent advancements have involved deep learning models. Abdukhamet<sup>9</sup> applied state-of-the-art deep learning models for landfill detection in satellite images of Shanghai district, treating the problem as an object detection task. Using models like RetinaNet<sup>10</sup> with DenseNet backbone<sup>11</sup>, he achieved 84.7 percent average precision in identifying waste sites. Similarly, Youme et al.<sup>12</sup> used drone images and a single shot detector object detection model to detect dumped waste. In 2021, Torres and Fraternali<sup>13</sup> presented a new solution for image classification on high-resolution red-green-blue (RGB) satellite and aerial imagery in Italy using the AerialWaste dataset. This dataset comprised images from the WorldView-3 satellite with a spatial resolution of approximately 30 cm, Google Earth images with a resolution of about 50 cm, and AGEA orthophotos with a resolution of approximately 20 cm. They employed a ResNet50 architecture enhanced with a feature pyramid network (FPN), achieving a precision of 0.88 and a recall of 0.87 on this dataset.

Besides imagery, GIS and structured data have also been used for waste detection. Whether in isolation, as shown by Jordá-Borrell et al.,<sup>14</sup> or in combination with remote sensing analysis, as done by Biotto et al.<sup>15</sup> and again by Silvestri and Omri<sup>16</sup>, socioeconomic data, land-use maps and geographic characteristics can help create predictive models for illegal landfill locations, revealing that such sites are often non-randomly distributed and concentrated in areas with certain predictive factors. Other notable efforts in detecting illegal debris dumping were carried out by UNDP Türkiye in collaboration with a private parafoil drone firm. They evaluated the effectiveness of parafoil drones in identifying the location and size of debris in the Kahramanmaraş Metropolitan Municipality, Türkiye. The study demonstrated that parafoil drones can serve as cost-effective and environmentally friendly alternatives to traditional remote sensing devices, such as regular drones, especially for monitoring larger areas<sup>17</sup>. These approaches can potentially improve the accuracy of existing remote sensing models.

---

<sup>9</sup> Shyngys Abdukhamet, "Landfill detection in satellite images using deep learning", PhD dissertation, Shanghai Jiao Tong University, 2019.

<sup>10</sup> Tsung-Yi Lin et al., "Focal loss for dense object detection", in *2017 IEEE International Conference on Computer Vision (ICCV)* (2017).

<sup>11</sup> Gao Huang et al., "Densely connected convolutional networks", in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).

<sup>12</sup> Ousmane Youme et al., "Deep learning and remote sensing: Detection of dumping waste using UAV", *Procedia Computer Science*, vol. 185 (2021), pp. 361–369.

<sup>13</sup> Rocio Nahime Torres and Piero Fraternali, "Learning to identify illegal landfills through scene classification in aerial images", *Remote Sensing*, vol. 13, No. 22 (2021), p. 4520.

<sup>14</sup> Rosa Jordá-Borrell, Francisca Ruiz-Rodríguez and Ángel Luis Lucendo-Monedero, "Factor analysis and geographic information system for determining probability areas of presence of illegal landfills", *Ecological Indicators*, vol. 37 (2014), pp. 151–160.

<sup>15</sup> Giancarlo Biotto et al., "GIS, multi-criteria and multi-factor spatial analysis for the probability assessment of the existence of illegal landfills", *International Journal of Geographical Information Science*, vol. 23 (2009), pp. 1233–1244.

<sup>16</sup> Sonia Silvestri and Mohamed Omri, "A method for the remote sensing identification of uncontrolled landfills: Formulation and validation", *International Journal of Remote Sensing*, vol. 2 (2008), pp. 975–989.

<sup>17</sup> Ceyda Alpay, "Revolutionizing debris tracking: Using parafoil drones in Türkiye's earthquake aftermath", *UNDP Türkiye* (United Nations Development Programme, 22 February 2024). Available at <https://undp.org/turkiye/blog/revolutionizing-debris-tracking-using-parafoil-drones-turkiyes-earthquake-aftermath>.

# 3 Key objectives

## 3.1 Research questions

The primary aim is to explore the use of machine learning to help illegal landfill monitoring in Guatemala.

The key objectives are twofold:

- Develop an effective machine learning model for illegal landfill monitoring in Guatemala.
- Utilize cost-efficient satellite imagery to implement the monitoring model.

For this, state-of-the-art approaches are prioritized with the possibility of using transfer learning, which implies using model knowledge from other geographical areas. The first objective is to determine which machine learning algorithms are capable of accurately detecting illegal landfill sites across diverse geographical and environmental conditions within Guatemala. The second objective is to explore the utilization of insights from other locations, such as Italy, within the Guatemalan context, while acknowledging the challenges associated with working with suboptimal data. The third objective is to find a balance between the quality, quantity and cost-effectiveness of aerial images employed in the monitoring process.

## 3.2 Key steps

To achieve the aforementioned objectives, several key steps have been identified. First, the development and refinement of machine learning models will be conducted using combined datasets from Italy's AerialWaste and ground truth data from Guatemala. Second, a thorough methodological exploration and literature review will be undertaken to enhance understanding of applicable methodologies. Third, a strategy will be implemented utilizing Google API images<sup>18</sup> as a primary data source, aimed at creating a scalable and cost-effective solution for illegal landfill monitoring.

## 3.3 Deliverables

The project aims to deliver several tangible outcomes. These include the development of an image classification model tailored for detecting dumping sites, the establishment of a sandbox environment for the validation and updating of models, the creation of a knowledge product derived from the project's findings, and the provision of technical documentation to facilitate the scaling up of proven models.

---

<sup>18</sup> Google, "Google Maps Static API". Available at <https://developers.google.com/maps/documentation/maps-static/overview> (accessed on 30 May 2024).

# 4 Data

## 4.1 AerialWaste dataset

The AerialWaste dataset<sup>19</sup> that was used to pre-train the classification model is an open-source Italian aerial dataset developed by the Politecnico di Milano (Polimi). The AerialWaste dataset is part of the PERIVALLON project, funded by the European Union H2020 Programme, which aims to protect the European territory from organized environmental crime through intelligent threat detection tools. It comprises 3,478 positive and 6,956 negative samples. This dataset emerges from a remote sensing data-collection campaign focusing on the Lombardy region in Italy. The dataset is composed of three different types of imagery data:

- **AGEA orthophotos:** Images from an airborne campaign by the Italian Agriculture Development Agency, captured in 2018 over Lombardy, Italy. These images have a high spatial resolution of approximately 20 cm and are presented in a standard size of about 1000 × 1000 pixels.
- **WorldView-3 satellite (2021):** Images collected offer a slightly lower spatial resolution of around 30 cm, with each image measuring approximately 700 × 700 pixels.
- **Google Earth imagery:** Accessed via Google API, these images have a spatial resolution of about 50 cm and are upscaled to a uniform size of 1000 × 1000 pixels.

The dataset was created by experts from the Environmental Protection Agency of Lombardy across 105 municipalities. Negative samples, approximately double the count of positive ones, were randomly selected from the same municipalities. Figure 1 shows an example of a positive sample.

---

<sup>19</sup> Rocio Nahime Torres and Piero Fraternali, “AerialWaste dataset for landfill discovery in aerial and satellite images”, *Scientific Data*, vol. 10 (2023), p. 63.



**Figure 1:** An example of a positive sample from the AerialWaste dataset showing an aerial view of a waste dump location

## 4.2 Ground truth data (Guatemala)

The Government of Guatemala has compiled a comprehensive dumpsite dataset featuring data collected over several years across the entire country, as seen in Table 1. This process involves personnel visiting various sites, identifying dumpsites and recording their coordinates. The dataset we used encompasses around 2,000 locations collected in 2021 and 2022, formatted as a CSV file. Each dataset contains the following attributes: department, municipality, location description, latitude and longitude. To control the visual quality of the sample we perform a visual inspection, where high-resolution imagery from Google Maps is used to verify if the dumpsites are visible or not. The refined dataset includes 700 negative points (locations without dumpsites) and 340 positive points (locations with visible dumpsites). Within the positive points, the dataset further differentiates between evident and non-evident dumpsites based on the ease of visual detection. Specifically, there are 154 evident samples, where the dumpsites are clearly visible, and 186 non-evident samples, where it is difficult to discern the presence of a dumpsite.

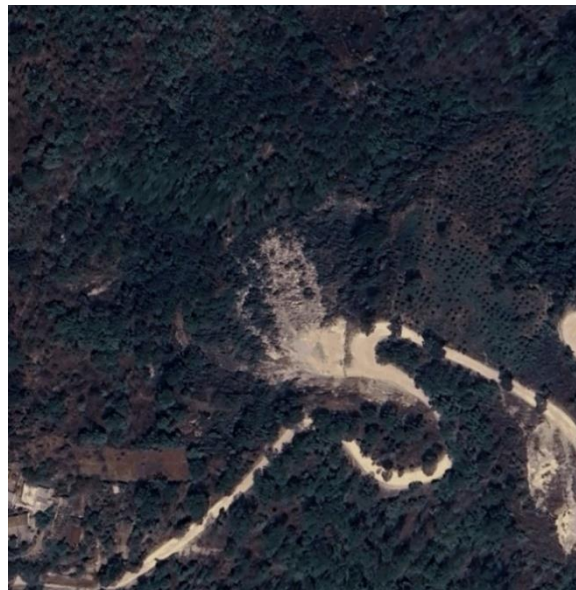
For illustrative examples of this visual inspection process, refer to the appendix of this paper. The appendix contains a series of images selected at random showing locations that were marked as positive for illegal dumpsites in 2022 but lacked visible evidence in the satellite imagery. They were thus not incorporated in our dataset.

**Table 1:** *Dumpsites shared by the Ministry of Environment and Natural Resources (MARN)*

Name	Source	Year	Data points
Illegal Dumpsites	MARN	2016	90
Illegal Dumpsites	MARN	2017	140
Illegal Dumpsites	MARN	2021	1,620
Illegal Dumpsites	MARN	2022	1,349

### 4.3 Image dataset (Guatemala)

Google Maps Static API<sup>20</sup> was used to collect a comprehensive set of aerial images for both positive and negative sample locations within Guatemala. Each image captured through the data pipeline has a size of 1000 pixels, with an approximate spatial resolution (ground sampling distance) of 30 cm and RGB bands. An example of a positive sample can be seen in Figure 2. Two distinct versions of the image dataset have been created. The first version focuses on ‘easy positives’, which comprises samples where the dumpsites are readily evident in the imagery. The second version includes a broader range of samples, encompassing both the evident and the more challenging, non-evident dumpsites. Although this version presents a greater challenge in terms of identifying dumpsites due to their less obvious nature, it offers a larger and more diverse set of samples.



**Figure 2:** An example of a positive sample from the Guatemala dataset

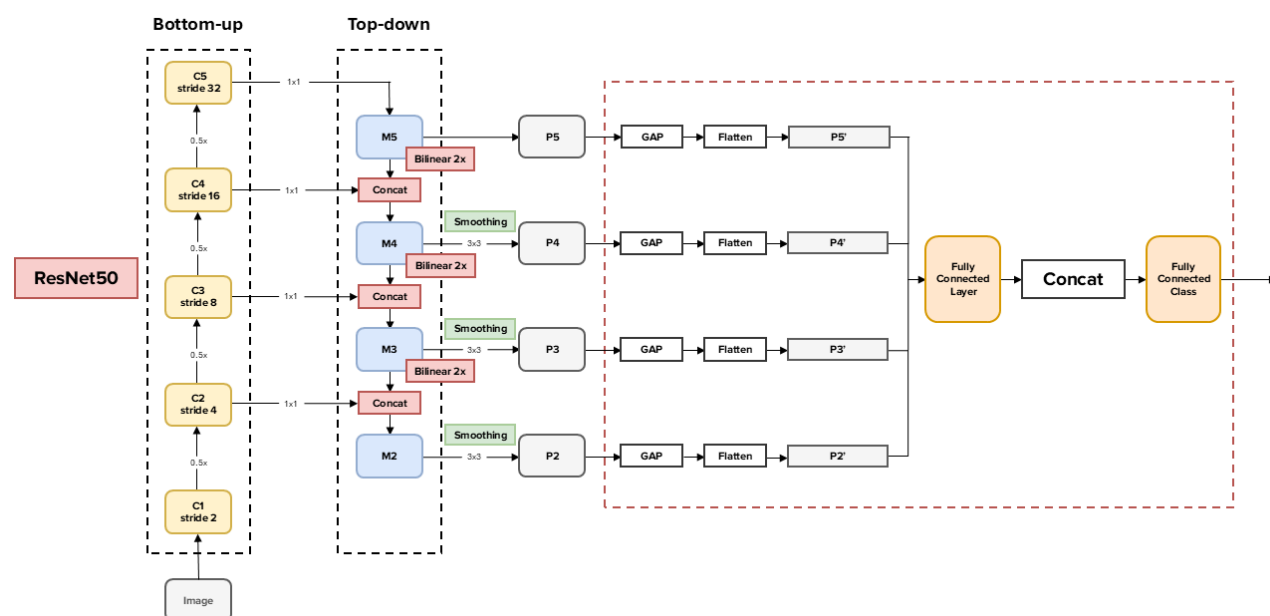
<sup>20</sup> Google, “Google Maps Static API”. Available at <https://developers.google.com/maps/documentation/maps-static/overview> (accessed on 30 May 2024).

# 5 Methodology

The methodology was strongly influenced by the previous achievements of Polimi<sup>21</sup>.

## 5.1 Model architecture

ResNet50 is a CNN architecture composed of 50 layers, designed to leverage residual connections for improved training and performance in image recognition tasks. The binary classifier exploits ResNet50 (initialized with ImageNet weights) as the network backbone and augments it with an FPN architecture.<sup>22</sup>



**Figure 3.** Example images and classification scores for illegal landfill detection. C1–C5 represent the different convolutional layers of the ResNet50 backbone; M2–M5 represent feature maps at different stages in the top-down pathway, typically part of a feature pyramid network (FPN); Concat is concatenation; P2–P5 are the output feature maps from each stage of the FPN after processing operations; GAP is global average pooling; Flatten converts the feature map into a one-dimensional vector for classification or further processing.

<sup>21</sup> Corrado Fasana and Samuele Pasini, "Learning to detect illegal landfills in aerial images with scarce labeling data", PhD dissertation, Politecnico di Milano, 2022. See also Rocio Nahime Torres and Piero Fraternali, "Learning to identify illegal landfills through scene classification in aerial images," *Remote Sensing*, vol. 13, No. 22 (2021), p. 4520.

<sup>22</sup> Rocio Nahime Torres and Piero Fraternali, "AerialWaste: A dataset for illegal landfill discovery in aerial images", *AerialWaste* (2022). Available at <https://aerialwaste.org/>.

The FPN architecture<sup>23</sup> improves performance in image classification when different scales are considered. The FPN exploits different semantic levels by combining low-resolution semantically strong features with high-resolution semantically weaker ones. This is realized by complementing the bottom-up feature extraction path typical of CNNs with a top-down path that merges features at multiple levels.

## 5.2 Transfer learning on Guatemala dataset

To train the model for the Guatemala data, the ResNet50 parameters are frozen, in a process called transfer learning. Transfer learning is a machine learning technique where a model developed for a specific task is reused as the starting point for a model on a second task. Initially, the model uses ResNet50 as the backbone, which is a deep CNN pre-trained on ImageNet, a large visual database used for image recognition tasks. ImageNet training gives the model a solid foundation in recognizing various features in images, such as textures, shapes and patterns. The model is then trained on the AerialWaste dataset for identifying waste dumps, learning specific features like size and texture. In the final stage, it adapts to Guatemalan conditions using our dataset, retaining the core structure and knowledge from previous training.

## 5.3 Dataset sampling

Our dataset is divided into two main subsets: the training set and the validation set. We allocate 80 percent of the data to the training set and 20 percent to the validation set. This split is achieved through random sampling, ensuring that each subset is a representative sample of the overall dataset. The limited number of positive samples (340) poses a significant challenge, making the training process prone to overfitting. Given the small size of the training set, removing a higher proportion of positive samples for validation would exacerbate the problem. We thus do not use a test set for the validation process. The training pipeline includes transformations applied directly to the input images. These transformations include random vertical and horizontal flipping. They are applied on the training data, allowing for a more robust training process by simulating various real-world conditions that the model might encounter when analysing aerial images. This approach helps in improving the generalization capability of the model when applied to new cases.

## 5.4 Validation metrics: Precision, recall and F1 score

Given the imbalance in our dataset, with a predominance of negative samples over positive ones, overall accuracy can be misleading. To effectively assess the performance of our model, we use precision, recall and F1 score. These metrics are particularly suitable for datasets with an imbalanced class distribution,

---

<sup>23</sup> Tsung-Yi Lin et al., “Feature pyramid networks for object detection”, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 936–944.

providing a more accurate reflection of the model's ability to correctly identify positive cases (illegal dumpsites) amid a large number of negatives (non-dumpsites).

**Precision** quantifies the accuracy of the model in predicting positive samples. In essence, it measures the proportion of correctly predicted positive observations (true positives) to the total predicted positives (true positives and false positives). This metric reflects the model's capability to correctly identify actual dumpsites, minimizing the false alarms (false positives). High precision implies that the model is reliable in its positive predictions. The precision score is calculated as follows:

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}} \quad (1)$$

**Recall**, also known as sensitivity, assesses the model's ability to identify all relevant instances of illegal dumpsites. It calculates the ratio of correctly predicted positive observations (true positives) to all positive observations (true positives and false negatives). This metric is vital for our study as it indicates how well the model can detect dumpsites, even those that are less evident or harder to discern. The recall score is calculated as follows:

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}} \quad (2)$$

**F1 score** provides a balance between precision and recall, offering a single score that encapsulates both precision and recall. The F1 score is calculated as follows:

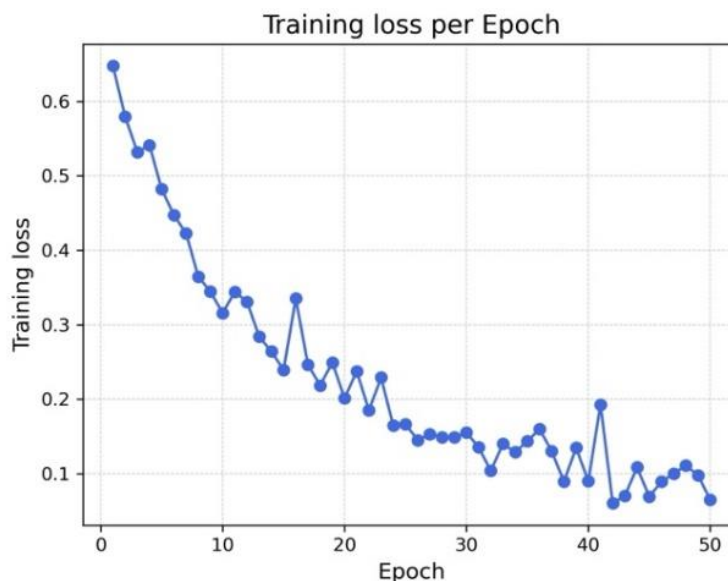
$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

## 6 Results

The progression of the machine learning model's performance over the course of training and validation phases is depicted in Figures 4 and 5, respectively. These figures provide valuable insights into the learning trajectory and the generalization capabilities of the model.

As illustrated in Figure 4, the training loss consistently decreased across epochs, with minor fluctuations indicative of the model's adaptation to the complexity of the dataset. Notably, by the final epoch, encompassing 10,000 iterations, the training loss significantly reduced to around 0.1, signifying effective learning from the training dataset and increasing prediction accuracy with epoch progression.

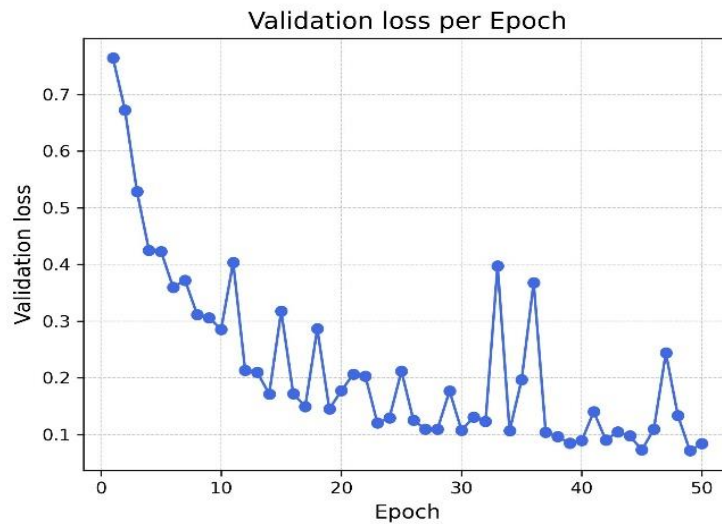
In the realm of machine learning, training loss and epoch are key terms used to gauge a model's learning progress over time. Training loss is depicted on the y axis and measures how far off a model's predictions are from the actual targets. It is an indication of the error rate; lower values indicate the model is making more accurate predictions. The epoch is shown on the x axis and counts how many times the model has gone through the entire dataset during training. Each epoch offers the model an opportunity to learn from its errors and improve its accuracy.



**Figure 4:** Training loss per epoch

As shown in Figure 5, the validation loss, starting at a higher initial value of around 0.8, followed a downward trend similar to the training loss, albeit with more pronounced spikes. These spikes are indicative of the challenges faced by the model in generalizing to unseen data, suggesting moments of potential overfitting. Despite this, the overall downward trend in validation loss, culminating in a value close to the training loss by the end of the epochs, is a positive indicator of the model's generalization ability and the quality of the

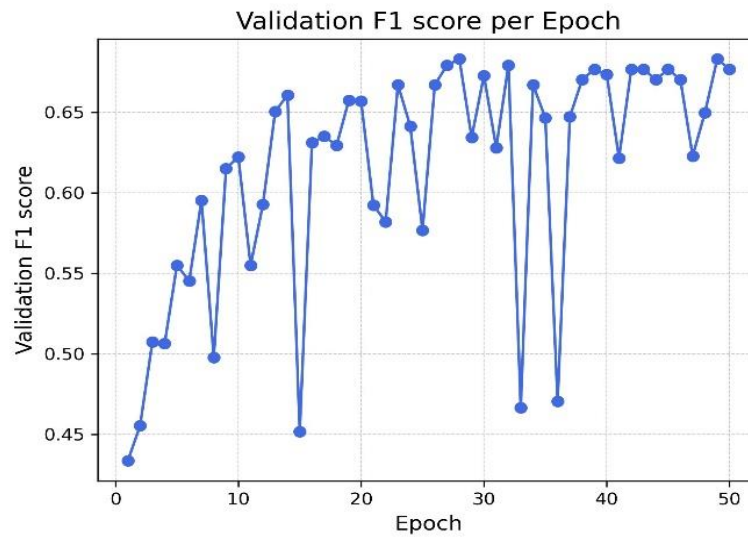
dataset. This convergence of training and validation loss values suggests that the model can also adapt to the dataset unbalance without significant overfitting or too much of a class imbalance bias.



**Figure 5:** Validation loss per epoch

The evaluation of the validation dataset metrics, as shown in Figures 6, 7 and 8, indicates the ability of the model to predict dumpsites with more accuracy, despite high variation. Precision and recall (Figures 7 and 8) fluctuated significantly, ranging from around 0.50 to 0.70 after a few epochs. The binary F1 score (Figure 6) combined these aspects, with values between 0.45 and 0.68. Despite these fluctuations, the general upward trend in these metrics suggests that the model quickly develops generalizing abilities and is very promising.

The analysis of the model's performance with the 'easy' dataset, which only consists of 154 dumpsite images, presents interesting insights. The limited number of positive samples primarily contributed to a significant degree of overfitting during the training. This overfitting is evident from the trends observed in the training loss and the highest F1 score, which remained relatively low, peaking at just around 0.36. After the initial epochs, the validation loss began to increase, reflecting the model's struggle to generalize beyond the training data. This issue of overfitting is starkly contrasted with the results obtained when training on the full dataset, which includes 340 positive images, encompassing both easy and hard positive samples. The inclusion of a larger number of positive samples, even the challenging ones, significantly allows the model to learn and generalize.



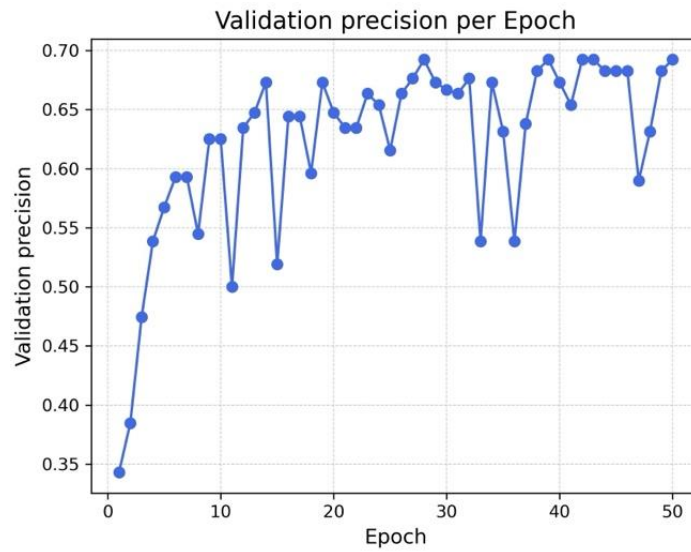
**Figure 6:** Validation binary F1 score

## 6.1 Final model results

The final model, trained on a combination of the AerialWaste dataset and the ground truth data from Guatemala, yielded promising results. The validation dataset consisted of 208 samples, with 140 negatives and 68 positives. We selected the model at epoch 48, which showed the best results in terms of validation metrics. The metrics achieved were:

- **Recall:** 0.68 – indicating that the model correctly identified 68 percent of actual illegal dumpsites.
- **Precision:** 0.68 – indicating that 68 percent of the locations identified by the model as illegal dumpsites were indeed such.
- **F1 score:** 0.68 – calculated as the harmonic mean of precision and recall.

The confusion matrix for the validation dataset is presented in Table 2. This matrix offers a detailed view of the model's performance, showing the number of correct and incorrect predictions for both classes.



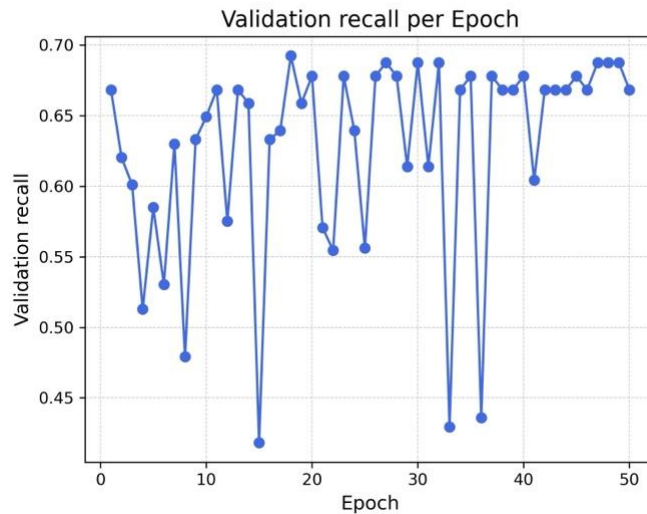
**Figure 7:** Validation binary precision

**Table 2:** Confusion matrix on the validation dataset (final model)

		Predicted	
		Negative	Positive
Actual	Negative	118	22
	Positive	21	47

## 7 Discussion

The initial results of our machine learning model for detecting illegal dumpsites in Guatemala show promising trends, especially in terms of the generalization capabilities indicated by the validation dataset metrics. However, several key considerations must be addressed to enhance the model's robustness and applicability.



**Figure 8:** Validation binary recall

In our methodology, we made a deliberate choice not to include a distinct test set due to the limited number of positive samples available. Creating a test set could lead to even stronger overfitting. This decision, while necessary under the current circumstances, does pose constraints on our ability to thoroughly assess the model's performance in real-world scenarios. The test set should ideally not merely be a randomized selection from the overall dataset; in the context of GIS and deep learning applications, it is often more effective to select the test set from a distinct geographical distribution. This separation, usually achieved by dividing the dataset along a specific longitudinal or latitudinal coordinate, ensures that the model is evaluated on geographically diverse samples. Such an approach is crucial for assessing the model's capacity to generalize across varied environmental conditions and characteristics of dumpsites, thereby mitigating the risk of overfitting to specific regional features.

## 8 Future work

Going forward, it would be beneficial to expand our dataset, particularly in terms of positive samples. This expansion would enable the creation of a distinct test set, providing a more comprehensive and realistic evaluation of the model's real-world applicability and effectiveness.

To expand the number of positives in our dataset, we could try to find and review other dumpsite datasets in Guatemala. We could also employ more robust labelling processes, such as using tools like Computer Vision Annotation Tool (CVAT) to enrich our dataset. Furthermore, incorporating inter-annotator agreement assessments would be advantageous in evaluating the quality of the imagery data. Inter-annotator agreement helps in determining the consistency and reliability of annotations provided by different individuals, thereby ensuring that the data used for training and testing the model is of high quality and

reliable. Manual interpretations and annotations of aerial images by interpreters, using tools like Google Earth, offer another straightforward approach. Expanding the size of the dataset is crucial for building a test set that provides a more realistic evaluation of the model's performance. Additionally, addressing potential long-term issues like data shift due to seasonal changes is important.

While increasing the number of positive samples has shown improvements in current image classification approaches, exploring alternative architectures could potentially yield even better results. For instance, investigating object detection models like You Only Look Once (YOLO) with transfer learning could provide fresh insights and enhanced accuracy in identifying illegal dumpsites. Moreover, leveraging advanced models such as ResNet-101, a residual network with up to 152 layers known for its depth and efficiency compared to earlier architectures like Visual Geometry Group (VGG), could further improve performance.<sup>24</sup> Another promising option is CrossFormer++,<sup>25</sup> a pre-trained model tailored for remote sensing with 2.4 billion parameters, demonstrating its potential in this domain. Additionally, experimenting with transformer-based architectures such as SWIN Transformer (SWiNT),<sup>26</sup> Vision Transformer (ViT),<sup>27</sup> Vision Transformer Advanced by Exploring Intrinsic Inductive Bias (ViTAE)<sup>28</sup> and Vision Transformer Advanced by Exploring Inductive Bias for Image Recognition and Beyond (ViTAEv2),<sup>29</sup> which are adaptable to various tasks including low-resolution images, could offer valuable alternatives.

With a more robust and extensive dataset, we can envision deploying this model across Guatemala. Utilizing large databases such as MillionAID30,<sup>30</sup> a comprehensive benchmark dataset with one million instances for remote sensing scene classification across 51 semantic categories tailored to land-use standards, would significantly enhance our capabilities. This broader application would enable thorough risk analysis and the establishment of a robust operational pipeline. Such expansion efforts would make substantial contributions to national environmental protection and the monitoring of illegal dumpsites. In the longer term, establishing a human-in-the-loop monitoring process could be highly beneficial. This approach involves utilizing the model to identify potential risk locations, which are then reviewed by domain experts. Beyond enhancing governmental effectiveness in detecting illegal dumpsites, insights from these reviews can continuously refine the model through iterative training processes. This iterative approach addresses challenges such as model decay and concept shift, ensuring sustained relevance and accuracy over time.

---

<sup>24</sup> Kaiming He et al., "Deep residual learning for image recognition", *arXiv:1512.03385 [cs.CV]* (10 December 2015). Available at <https://arxiv.org/abs/1512.03385>.

<sup>25</sup> Keumgang Cha et al., "A billion-scale foundation model for remote sensing images", *arXiv:2304.05215 [cs.CV]* (11 April 2023). Available at <https://arxiv.org/abs/2304.05215>.

<sup>26</sup> Ze Liu et al., "SWIN Transformer: Hierarchical vision transformer using shifted windows", *arXiv:2103.14030 [cs.CV]* (25 March 2021). Available at <https://arxiv.org/abs/2103.14030v2>.

<sup>27</sup> Alexey Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale", *arXiv:2010.11929 [cs.CV]* (22 October 2020). Available at: <https://arxiv.org/abs/2010.11929v2>.

<sup>28</sup> Yufei Xu et al., "ViTAE: Vision Transformer Advanced by Exploring Intrinsic Inductive Bias", *arXiv:2106.03348 [cs.CV]* (7 June 2021). Available at <https://arxiv.org/abs/2106.03348v4>.

<sup>29</sup> Qiming Zhang et al., "ViTAEv2: Vision Transformer Advanced by Exploring Inductive Bias for Image Recognition and Beyond", *International Journal of Computer Vision*, vol. 131 (2023), pp. 1141–1162.

<sup>30</sup> Yang Long et al., "On creating benchmark dataset for aerial image interpretation: Reviews, guidances and Million-AID", *Papers with Code* (22 June 2020). Available at <https://paperswithcode.com/paper/dirs-on-creating-benchmark-datasets-for>.

Alternatively, building an application programming interface (API) for the model could allow various stakeholders, including governmental and non-governmental organizations, to access and utilize the tool for monitoring and intervention purposes. This step could contribute to a more collaborative approach in tackling the challenge of illegal dumpsites. The exploration of alternative remote sensing methods, such as parafoil drones, could help to enhance the quality of existing dumpsite images further. By doing so, we can ensure that the data collected is both accurate and comprehensive, thereby supporting more effective waste management strategies. Additionally, integrating these advanced technologies and methodologies can provide more detailed and actionable insights, ultimately contributing to the national efforts in environmental protection and illegal dumpsite monitoring.

## 9 Conclusion

In conclusion, this project successfully demonstrates the application of machine learning and GIS techniques for identifying illegal dumpsites in Guatemala, providing state-of-the-art methods for shaping and monitoring environmental policies. Using state-of-the-art computer vision methods both in architectures (ResNet50 and FPN) and training methodology (transfer learning), we have developed a binary classifier for satellite imagery analysis able to generalize to new areas in Guatemala. Although promising, the results highlight the need for a more diverse and extensive dataset to enhance model accuracy and reliability, and exploring alternative model architectures. Also, a future key area of improvement will be to enhance the monitoring process by incorporating a human-in-the-loop system for monitoring and continuous model refinement. Finally, the development of an API for wider interoperability is an important step forward. This project significantly contributes to environmental protection efforts in Guatemala and offers a state-of-the-art scalable method for other regions facing similar challenges.

**Appendix:** Locations marked as positive in 2022 for illegal dumpsites but lacking visible evidence in satellite imagery



# Bibliography

- Abbott, Jeff. Guatemala's rivers of garbage. *The Progressive*, 16 August 2023.
- Abdukhamet, Shynggys. Landfill detection in satellite images using deep learning. PhD dissertation, Shanghai Jiao Tong University, 2019.
- Alpay, Ceyda. Revolutionizing debris tracking: Using parafoil drones in Türkiye's earthquake aftermath. *UNDP Türkiye*, United Nations Development Programme, 22 February 2024. Available at <https://undp.org/turkiye/blog/revolutionizing-debris-tracking-using-parafoil-drones-turkiyes-earthquake-aftermath>.
- Angelino, Cesario Vincenzo, et al. A case study on the detection of illegal dumps with GIS and remote sensing images. In *Earth Resources and Environmental Remote Sensing/GIS Applications IX*, vol. 10790. SPIE Remote Sensing, 2018.
- Biotto, Giancarlo, et al. GIS, multi-criteria and multi-factor spatial analysis for the probability assessment of the existence of illegal landfills. *International Journal of Geographical Information Science*, vol. 23 (2009), pp. 1233–1244.
- Cha, Keumgang, et al. A billion-scale foundation model for remote sensing images. *arXiv:2304.05215 [cs.CV]*, 11 April 2023. Available at <https://arxiv.org/abs/2304.05215>.
- Dosovitskiy, Alexey, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv:2010.11929 [cs.CV]*, 22 October 2020. Available at: <https://arxiv.org/abs/2010.11929v2>.
- Fasana, Corrado, and Samuele Pasini. Learning to detect illegal landfills in aerial images with scarce labeling data. PhD dissertation, Politecnico di Milano, 2022.

- Garofalo, Donald, and Frank J. Wobber. Solid waste and remote sensing. *Photogrammetric Engineering*, vol. 40 (1974), pp. 45–59.
- He, Kaiming, et al. Deep residual learning for image recognition. *arXiv:1512.03385 [cs.CV]*, 10 December 2015. Available at <https://arxiv.org/abs/1512.03385>.
- Huang, Gao, et al. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- Jordá-Borrell, Rosa, Francisca Ruiz-Rodríguez and Ángel Luis Lucendo-Monedero. Factor analysis and geographic information system for determining probability areas of presence of illegal landfills. *Ecological Indicators*, vol. 37 (2014), pp. 151–160.
- Lin, Tsung-Yi, et al. Feature pyramid networks for object detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- Lin, Tsung-Yi, et al. Focal loss for dense object detection. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017.
- Liu, Ze, et al. SWIN Transformer: Hierarchical vision transformer using shifted windows. *arXiv:2103.14030 [cs.CV]*, 25 March 2021. Available at <https://arxiv.org/abs/2103.14030v2>.
- Long, Yang, et al. On creating benchmark dataset for aerial image interpretation: Reviews, guidances and Million-AID. *Papers with Code*, 22 June 2020. Available at <https://paperswithcode.com/paper/dirs-on-creating-benchmark-datasets-for>.
- Lyon, John Grimson. Use of maps, aerial photographs, and other remote sensor data for practical evaluations of hazardous waste sites. *Photogrammetric Engineering and Remote Sensing*, vol. 53 (1987), pp. 515–519.

- Selani, Lungile. Mapping illegal dumping using a high resolution remote sensing image: Case study: Soweto Township in South Africa. PhD dissertation, University of the Witwatersrand, 2017.
- Silvestri, Sonia, and Mohamed Omri. A method for the remote sensing identification of uncontrolled landfills: Formulation and validation. *International Journal of Remote Sensing*, vol. 2 (2008), pp. 975–989.
- Torres, Rocio Nahime, and Piero Fraternali. AerialWaste dataset for landfill discovery in aerial and satellite images. *Scientific Data*, vol. 10 (2023), p. 63.
- Torres, Rocio Nahime, and Piero Fraternali. Learning to identify illegal landfills through scene classification in aerial images. *Remote Sensing*, vol. 13, No. 22 (2021), p. 4520.
- Xu, Yufei, et al. ViTAE: Vision Transformer Advanced by Exploring Intrinsic Inductive Bias. *arXiv:2106.03348 [cs.CV]*, 7 June 2021. Available at <https://arxiv.org/abs/2106.03348v4>.
- Youme, Ousmane, et al. Deep learning and remote sensing: Detection of dumping waste using UAV. *Procedia Computer Science*, vol. 185 (2021), pp. 361–369.
- Zhang, Qiming, et al. ViTAEv2: Vision Transformer Advanced by Exploring Inductive Bias for Image Recognition and Beyond. *International Journal of Computer Vision*, vol. 131 (2023).



Copyright © UNDP 2025 All rights reserved. The views expressed in this publication are those of the authors and do not necessarily represent those of the United Nations, including UNDP, or UN Member States.