

# LARGE MULTIMODAL MODELS ENABLE SCALABLE MONITORING OF AQUACULTURE PONDS

**Anmol Kabra<sup>1</sup>, Felipe Siqueira Pacheco<sup>2</sup>, Brendan Hogan<sup>1</sup>, Laura Greenstreet<sup>1</sup>, Joshua Fan<sup>1</sup>, Aaron Ferber<sup>1</sup>, Marta Ummus<sup>3</sup>, Alecsander Brito<sup>3</sup>, Alex Flecker<sup>2</sup>, Carla Gomes<sup>1</sup>**

<sup>1</sup>Department of Computer Science, Cornell University, Ithaca, NY 14850, USA

<sup>2</sup>Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14850, USA

<sup>3</sup>Empresa Brasileira de Pesquisa Agropecuária, Brasília, DF, Brazil

{ak2426, felipe.pacheco, bhr54, leg86, jy6, amf272}@cornell.edu  
{marta.ummus, alecsander.brito}@embrapa.br, {asf3, cpg5}@cornell.edu

## ABSTRACT

Aquaculture plays an increasingly important role in global food systems, and its rapid expansion, particularly in biodiverse regions such as the Amazon River basin, requires reliable and scalable spatial monitoring to ensure environmentally sustainable management. However, satellite-based monitoring of inland aquaculture systems has traditionally been constrained by the difficulty of detecting small and irregularly shaped aquaculture ponds whose spectral similarity to surrounding land uses and variable optical conditions complicate their identification in heterogeneous landscapes. Consequently, conventional classification approaches (e.g., Random Forest and CNNs) often fail to capture this spatial and contextual complexity. Here, we present a novel agentic framework, AISciVision-Aqua, which integrates Retrieval-Augmented Generation (RAG), large multimodal models (LMMs), and agentic interactive tools (e.g., zooming, panning, and predictive tools) to explicitly emulate human expert workflows and detect aquaculture ponds from satellite imagery. Our results demonstrate that AISciVision-Aqua consistently outperforms baseline classification methods, such as convolutional neural networks, achieving higher precision and recall while providing transparent reasoning transcripts that detail the model’s decision-making process. The AISciVision-Aqua’s interactive capabilities enable ecologists to validate, interrogate (in different languages), and iteratively refine predictions in real-time, fostering trust and adaptability in Artificial Intelligence (AI)-assisted environmental monitoring. Collectively, these findings demonstrate how LMM-driven, agentic reasoning can advance fine-scale environmental feature extraction from satellite imagery, offering a scalable and interpretable approach for mapping inland aquaculture. AISciVision-Aqua thus exemplifies next-generation remote sensing workflows that tightly integrate human expertise and agentic multimodal AI for scalable environmental monitoring and informed decision making.

## 1 INTRODUCTION

Inland aquaculture now accounts for more than half of global farmed fish production, and much of its recent expansion is occurring in tropical regions undergoing rapid land-use change, widespread biodiversity loss, and accelerating climate impacts, while these same regions often lack the high-quality data needed for effective monitoring. At the same time, in these regions, aquaculture holds significant potential to support sustainable food production, economic development, and food security while imposing comparatively lower environmental costs (FAO, 2024; Osmundsen et al., 2020). In the Amazon, aquaculture has expanded over the past four decades, but production remains consistently underreported, and reliable information on the location and extent of aquaculture ponds is still limited (Pacheco et al., 2025; McGrath et al., 2020). Yet, understanding the spatial distribution of aquaculture ponds is essential for assessing the environmental impacts of this activity, including greenhouse gas emissions, land-use changes, and potential impacts on biodiversity, as well as

evaluating its contribution to food production in the region. Improving spatial data on aquaculture is therefore a prerequisite for guiding sustainable development strategies in rapidly transforming landscapes, such as the Amazon.

Although mapping surface water from satellite imagery is a well-established remote sensing application, mapping aquaculture ponds presents a distinct challenge because their detection depends not only on spectral signatures but also on spatial and contextual features such as shape, arrangement, and proximity to infrastructure or other land uses. Therefore, conventional approaches, such as Random Forest classifiers and spectral-based algorithms, often struggle to distinguish aquaculture ponds from visually similar natural water bodies. This detection problem remains among the most challenging tasks in aquatic land-cover mapping. Globally, several efforts have addressed the challenge of aquaculture pond detection, including in regions such as Southeast Asia and parts of Brazil (Silva et al., 2024; Luo et al., 2022). Yet the Amazonian landscape further complicates these challenges due to its highly heterogeneous land cover and the wide variation in pond size, shape, and structure. For example, ponds are frequently built in river channels and consequently follow the natural river network, which makes them difficult to distinguish from rivers and floodplain channels in satellite imagery (Freitas et al., 2022). Given these complexities and the limited resolution of most freely available satellite data, accurate pond identification often requires extensive manual interpretation and validation, which are labor-intensive, expensive, and difficult to scale over large regions. These constraints underscore the need for approaches that integrate both spectral and spatial context features to enable accurate, scalable, and interpretable detection of aquaculture systems in the Amazon and similarly complex regions.

Recent advancements in artificial intelligence (AI), particularly in large multimodal models (LMMs) and agentic AI systems, have opened new avenues for domain-specific applications in environmental science and remote sensing (Gomes et al., 2019). LMMs are designed to process and integrate heterogeneous data modalities, such as imagery, text, and metadata, which enables joint representations that capture spatial and contextual relationships often missed by conventional spectral-based classifiers. In parallel, agentic AI systems operate at a higher level of abstraction by orchestrating tools, models, and iterative reasoning steps, thereby emulating expert analytical workflows such as spatial pattern recognition, cross-scale contextual reasoning, and hypothesis refinement. When combined with retrieval-augmented generation (RAG) to incorporate external knowledge and prior information, these components form a foundation for Agentic Multimodal Models (AMMs) that have the potential to deliver accurate, interpretable, and context-aware predictions for complex environmental features such as aquaculture ponds. However, despite growing recognition of the promise of LMMs and agentic systems for remote sensing, no operational frameworks or end-to-end workflows have, to our knowledge, been proposed that systematically integrate multimodal learning, agentic reasoning, retrieval mechanisms, and interpretability to address the core challenges of environmental feature extraction requiring the joint use of spectral, spatial, and contextual information.

Here, we introduce AISciVision-Aqua, a novel agentic framework that detects aquaculture ponds using high-resolution satellite imagery and AI agents. Our work pursues three complementary aims: (1) to demonstrate the efficacy of AI-enhanced remote sensing in detecting aquaculture systems in this region through autonomous, agentic reasoning and action selection, (2) to provide a transparent and interpretable decision-making process for ecological monitoring that enables human oversight, validation, and iterative refinement of model predictions, and (3) to explore the implications of this technology for sustainable aquaculture development and for advancing both remote sensing and AI toward reliable, real-world deployment in environmental science. This study lays the foundation for developing transparent, accurate, and transferable AI approaches for environmental monitoring, advancing the integration of expert knowledge and AI in remote sensing applications.

**Study Sites.** This study focuses on aquaculture in Rondônia, Brazil, a state located in the heart of the Amazon’s “arc of deforestation” (Figure 1), which has experienced rapid development in aquaculture since the early 2000s. Rondônia is characterized by a mosaic of natural forests, degraded pastures, and agricultural lands. Aquaculture ponds in Rondônia are often located near river networks or low-lying areas, making them distinguishable in high-resolution imagery. The interplay between aquaculture expansion and land-use dynamics makes this region a compelling case for applying advanced remote sensing techniques for detecting aquaculture ponds.

**Definition of “easy” and “hard” examples.** Satellite imagery-based detection of aquaculture ponds in the Amazon presents varying difficulty levels (Figure 2). “Easy examples” refer to aquaculture

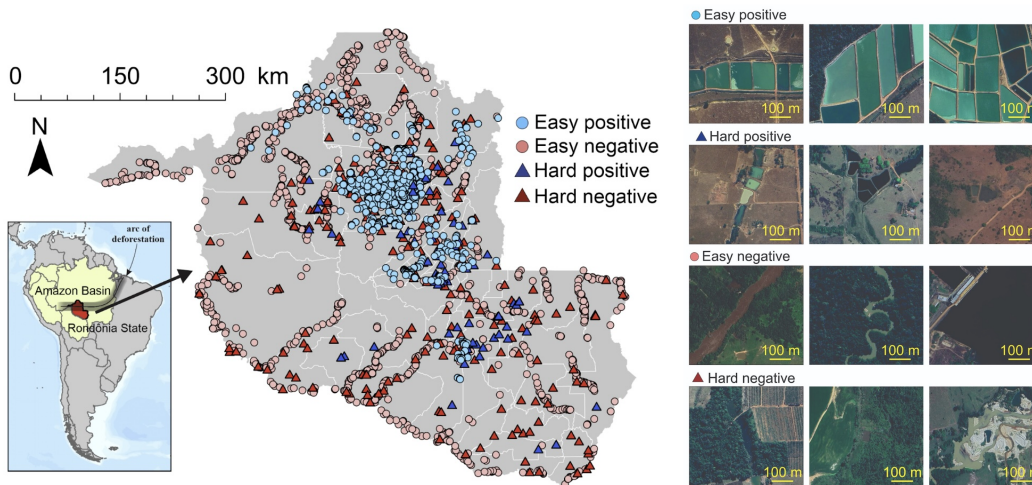


Figure 1: Geographic distribution of water body samples used for aquaculture pond detection in Rondônia State, Brazil. Symbols represent georeferenced examples classified as: easy positive, aquaculture ponds clearly identifiable (light blue circles); easy negative, natural water bodies easily distinguished from aquaculture ponds (pink circles); hard positive, aquaculture ponds with irregular shapes or visually complex surroundings (dark blue triangles); and hard negative, non-aquaculture features visually similar to ponds, such as clearings or natural wetlands (dark red triangles).

Figure 2: Representative examples of easy and hard positive (aquaculture ponds) and negative (non-aquaculture ponds) samples in Rondônia, Brazil. Easy examples show a clear visual distinction between classes, whereas hard examples exhibit subtle or ambiguous features that challenge automated classification.

ponds with regular geometric shapes, clearly defined boundaries, and high visual contrast, making them readily detectable. “Hard examples” feature smaller size, irregular shapes, subtle boundaries, and low contrast, leading to detection failures. We recognize these levels to design our AISciVision-Aqua framework, which handles the complexity typical of Amazonian aquaculture systems.

## 2 METHODOLOGY

This work introduces AISciVision-Aqua, an agentic AI framework for detecting aquaculture ponds in satellite imagery from Rondônia, Brazil. The framework (Figure 3) comprises three stages: initial AI-assisted water detection (Stage 1), AI-assisted labeled data organization (Stage 2), and AI-driven waterbody classification (Stage 3). The final stage integrates agentic AI through an LMM with Visual Retrieval-Augmented Generation (VisRAG) and domain-specific interactive tools for spatial interaction and contextual reasoning, enabling autonomous iterative refinement of predictions in a manner that mirrors expert workflows. See Appendix A for implementation details.

## 3 RESULTS

We evaluate on 3700 easy examples (1900 fish pond, 1800 non-fish pond) and 1000 hard examples (100 fish pond, 200 non-fish pond). We describe the data collection procedure in Appendix A, with “easy” and “hard” dataset breakdown in Table 1.

AISciVision-Aqua using the best performing LMM (Gemini-2.5-Flash) substantially enhanced aquaculture pond detection performance compared to traditional baseline methods (Figure 4). The improvement in F1 score was primarily driven by a large increase in recall, indicating that **AISciVision-Aqua is much better at minimizing missed aquaculture ponds (false negatives)**. The AISciVision-Aqua framework supports all LMMs, from GPT-5 to Llama-3.2-90B. Not all perform equally in classifying ponds (b)—Gemini-2.5-Flash is the best—but the framework provides a general performance boost. Furthermore, the full framework configuration (LMM+VisRAG+Tools) performs better than combination of the parts (c). In Appendix B we stratify performance on “easy” and “hard” examples, and illustrate classification results of the framework.

## 4 CONCLUSION

AISciVision-Aqua introduces a powerful, interpretable, and scalable interactive agentic AI framework for aquaculture monitoring in complex tropical landscapes. By combining retrieval-augmented reasoning, interactive spatial tools, and LMMs, the system effectively mimics human expert workflows and significantly improves the accuracy and transparency of aquaculture pond detection. This

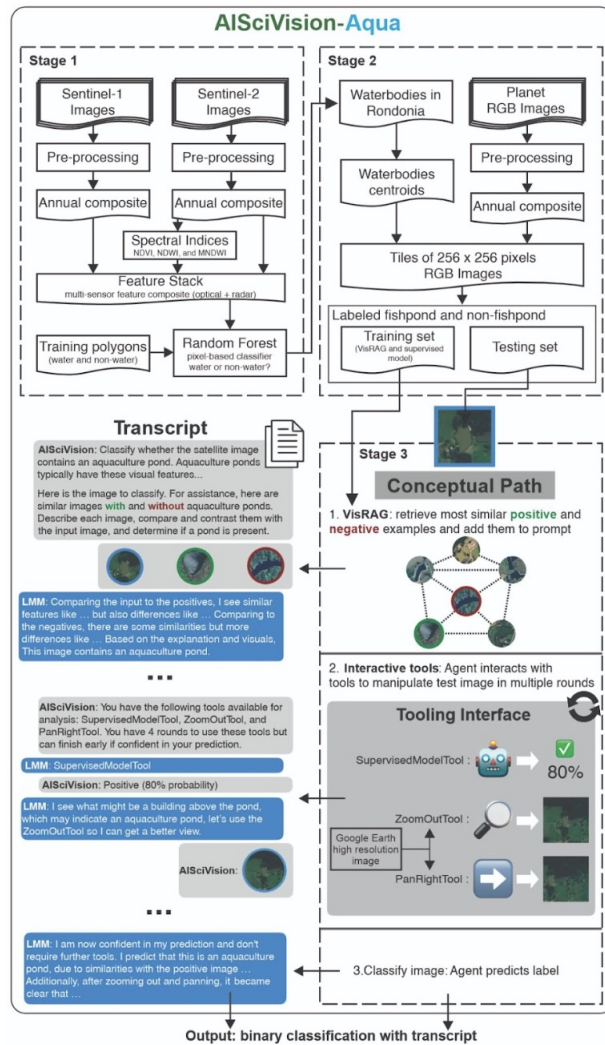


Figure 3: Workflow of the AISciVision-Aqua framework for aquaculture pond classification. Waterbodies are first identified using a Random Forest model trained on satellite imagery. These candidate sites are then analyzed by the AISciVision-Aqua agent, which (1) retrieves visually similar and dissimilar examples using a retrieval-augmented strategy, and (2) interacts with a set of tools (e.g., zoom, pan, and machine learning-based predictors) via a large multimodal model (LMM). The agent outputs a binary classification (aquaculture pond or not) along with a transcript that records the step-by-step reasoning process. See Appendix A for details.

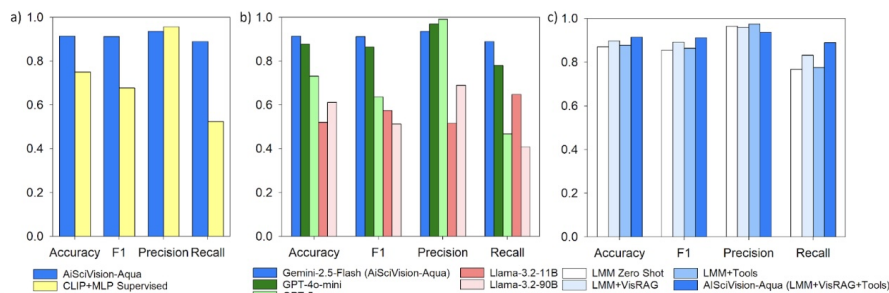


Figure 4: Comparative performance of AISciVision-Aqua and model variants across four evaluation metrics using the full dataset. (a) AISciVision-Aqua consistently outperforms the supervised approach (CLIP+MLP), particularly in recall. (b) Among various backbone LMM choices for the framework, Gemini-2.5-Flash achieved the most balanced performance across metrics. (c) The full framework configuration (LMM+VisRAG+Tools) delivered the strongest and most consistent performance, underscoring the complementary contributions of retrieval-augmented reasoning and tool-based interaction.

approach advances current practices in remote sensing, which have traditionally relied on static classification models such as Random Forests or CNNs, often limited by their inability to handle irregular features and explain classification decisions. AISciVision-Aqua not only advances aquaculture mapping but also sets the stage for the next generation of interactive agentic remote sensing tools that integrate human expertise and AI in environmental monitoring and food system planning.

#### ACKNOWLEDGEMENTS

This project is partially supported by the National Science Foundation (NSF); the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a program of Schmidt Sciences, LLC; the National Institute of Food and Agriculture (USDA/NIFA); the Air Force Office of Scientific Research (AFOSR); Toyota Research Institute (TRI); and a Schmidt AI2050 Senior Fellowship, a Schmidt Sciences program.

#### REFERENCES

- FAO. The State of World Fisheries and Aquaculture 2024 – Blue Transformation in action. 2024.
- Carlos EC Freitas, Marcos de Almeida Mereles, Diego Valente Pereira, Flavia Siqueira-Souza, Lawrence Hurd, James Kahn, Gilberto Morais, and Raniera Garcez Costa Sousa. Death by a thousand cuts: Small local dams can produce large regional impacts in the Brazilian Legal Amazon. *Environmental Science & Policy*, 136:447–452, 2022. ISSN 1462-9011. doi: <https://doi.org/10.1016/j.envsci.2022.07.013>. URL <https://www.sciencedirect.com/science/article/pii/S1462901122002209>.
- Carla Gomes, Thomas Dietterich, Christopher Barrett, Jon Conrad, Bistra Dilkina, Stefano Ermon, Fei Fang, Andrew Farnsworth, Alan Fern, Xiaoli Fern, Daniel Fink, Douglas Fisher, Alexander Flecker, Daniel Freund, Angela Fuller, John Gregoire, John Hopcroft, Steve Kelling, Zico Kolter, Warren Powell, Nicole Sintov, John Selker, Bart Selman, Daniel Sheldon, David Shmoys, Milind Tambe, Weng-Keen Wong, Christopher Wood, Xiaojian Wu, Yexiang Xue, Amulya Yadav, Abdul-Aziz Yakubu, and Mary Lou Zeeman. Computational sustainability: computing for a better world and a sustainable future. *Commun. ACM*, 62(9):56–65, August 2019. ISSN 0001-0782. doi: 10.1145/3339399. URL <https://doi.org/10.1145/3339399>.
- Juhua Luo, Zhe Sun, Lirong Lu, Zhuyang Xiong, Lipeng Cui, and Zhigang Mao. Rapid expansion of coastal aquaculture ponds in Southeast Asia: Patterns, drivers and impacts. *Journal of Environmental Management*, 315:115100, 2022. ISSN 0301-4797. doi: <https://doi.org/10.1016/j.jenvman.2022.115100>. URL <https://www.sciencedirect.com/science/article/pii/S0301479722006739>.
- DG McGrath, L Castello, M Brabo, D Nepstad, S Da Gama, B Forsberg, E Mendoza, G Estupinan, M Ruffino, A Ribeiro, OT Almeida, AJ Bentes, and Chan C. Can fish drive development of the Amazon bioeconomy. *Policy brief for The Earth Innovation Institute*, 2020. URL <https://earthinnovation.org/Pub/21189/can-fish-drive-development-of-the-amazon-bioeconomy>.
- Tonje C. Osmundsen, Vilde S. Amundsen, Karen A. Alexander, Frank Asche, Jennifer Bailey, Bengt Finstad, Marit Schei Olsen, Klaudia Hernández, and Hugo Salgado. The operationalisation of sustainability: Sustainable aquaculture production as defined by certification schemes. *Global Environmental Change*, 60:102025, 2020. ISSN 0959-3780. doi: <https://doi.org/10.1016/j.gloenvcha.2019.102025>. URL <https://www.sciencedirect.com/science/article/pii/S0959378019304595>.
- Felipe S Pacheco, Sebastian A Heilpern, Claire DiLeo, Rafael M Almeida, Suresh A Sethi, Marcela Miranda, Nicholas Ray, Nathan O Barros, Jucilene Cavali, Carolina Costa, Carolina R Doria, Joshua Fan, Kathryn J Fiorella, Bruce R Forsberg, Marcelo Gomes, Laura Greenstreet, Meredith Holgerson, David McGrath, Peter B McIntyre, Patricia Moraes-Valenti, Ilce Oliveira, Jean P Ometto, Fabio Roland, Adry Trindade, Marta E Ummus, Wagner C Valenti, Xiangtao Xu, Carla P Gomes, and Alexander S Flecker. Towards sustainable aquaculture in the amazon. *Nature Sustainability*, 8(3):234–244, 2025.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Bruno Aparecido Silva, Marta Eichemberger Ummus, Ericson Hideki Hayakawa, Altair Bennert, Marcos Adami, Carolina Balera Trombini, Aldi Feiden, Kennedy Leocadio Vasco, and Alexander Gomes Brito. Mapping aquaculture in inland continental areas of brazil using machine learning on the google earth engine. *Remote Sensing Applications: Society and Environment*, 36: 101391, 2024. ISSN 2352-9385. doi: <https://doi.org/10.1016/j.rsase.2024.101391>. URL <https://www.sciencedirect.com/science/article/pii/S2352938524002556>.

## A DETAILED METHODOLOGY

### A.1 WATERBODY DETECTION (STAGE 1)

The first stage of the AISciVision-Aqua framework focuses on identifying all surface waterbodies across the state of Rondônia to generate the input base for subsequent classification (Figure 3). The workflow was implemented in the Google Earth Engine (GEE) cloud-computing platform. For each year (2019–2021), we built annual multi-sensor composites using Sentinel-2 MultiSpectral Instrument (MSI) optical imagery and Sentinel-1 C-band Synthetic Aperture Radar (SAR) data. For optical data, we used the Sentinel-2 Level-2A surface reflectance product, which includes atmospheric correction. All available images for each year were filtered for scene cloudiness ( $< 60\%$ ) and processed using a combined cloud and cloud-shadow masking approach based on the QA60 bitmask (cloud and cirrus flags) and the Scene Classification Layer (SCL), removing cloud, cloud shadow, and cirrus classes. Reflectance values were scaled to surface reflectance. Annual optical composites were generated using the median of all valid observations. Three spectral indices were computed from the median Sentinel-2 composite and added as predictor layers:

- $NDVI = B8 - B4 / B8 + B4$
- $NDWI = B3 - B8 / B3 + B8$
- $mNDWI = B3 - B11 / B3 + B11$

where NDVI is the Normalized Difference Vegetation Index, NDWI is the Normalized Difference Water Index, and mNDWI is the modified Normalized Difference Water Index. B3 (green, 560 nm), B4 (red, 665 nm), B8 (near-infrared, 842 nm) and B11 (short-wave infrared, 1610 nm) are Sentinel-2 MSI bands. We used the 10-m NIR band (B8) for NDVI and NDWI, and the SWIR-1 band (B11,  $\approx 1.6 \mu\text{m}$ ) for mNDWI.

Sentinel-1 Ground Range Detected (GRD) data were used to provide complementary radar information. We selected images acquired in Interferometric Wide (IW) swath mode, descending orbit, with 10-m spatial resolution, and extracted the VV and VH polarizations. For each year, median composites of VV and VH backscatter were generated. The final multi-sensor feature stack consisted of Sentinel-2 spectral bands (B2, B3, B4, B8, B11 and B12; 10–20 m), the three spectral indices (NDVI, NDWI and mNDWI), and Sentinel-1 VV and VH backscatter, resampled to a common 10m grid.

A random forest classifier was trained using manually delineated polygons representing water and non-water areas. The classifier was applied to the annual multi-sensor composites to produce a binary surface-water mask. This mask was subsequently used to delineate and extract individual candidate water objects, providing a comprehensive map of generic surface waterbodies across Rondônia that served as the foundation for identifying potential aquaculture ponds in later stages of the framework.

### A.2 LABELED DATA ORGANIZATION (STAGE 2)

The second stage focused on organizing and refining the labeled dataset used to train and validate the classification model. A validated database of aquaculture pond locations, containing more than 11,000 aquaculture ponds across the state of Rondônia, was provided by the Brazilian Agricultural Research Corporation (Embrapa). These data were manually verified by Embrapa experts through visual inspection of high-resolution imagery, ensuring high positional accuracy and consistent labeling across diverse landscape contexts. The expert-curated dataset served as ground truth, distinguishing aquaculture ponds from natural and artificial non-aquaculture ponds such as reservoirs, wetlands, and small natural lakes. Generic

Split	Class	Easy	Hard	Total
Full dataset	Positive	2006	208	2214
	Negative	1998	792	2790
	<b>Total</b>	<b>4004</b>	<b>1000</b>	<b>5004</b>
Train split	Positive	106	108	214
	Negative	198	592	790
	<b>Total</b>	<b>304</b>	<b>700</b>	<b>1004</b>
Test split	Positive	1900	100	2000
	Negative	1800	200	2000
	<b>Total</b>	<b>3700</b>	<b>300</b>	<b>4000</b>

Table 1: Summary of the labeled dataset for aquaculture pond detection in Rondônia, Brazil. Samples are classified into easy and hard cases, with positive (fish pond) and negative (non-fish pond) examples.

waterbodies identified in Stage 1 were intersected with the Embrapa reference database to assign aquaculture pond or non-aquaculture pond labels. To construct the final dataset, a sample of 5,004 labeled waterbodies was selected, comprising 2,214 positive (aquaculture pond) and 2,790 negative (non-aquaculture pond) examples. The dataset was divided into training (1004 samples) and testing (4,000 samples) subsets (Table 1). High-resolution PlanetScope RGB imagery corresponding to each waterbody was pre-processed, mosaicked into annual composites, and subdivided into  $256 \times 256$  pixel tiles centered on the labeled centroids. Each tile represented an independent observation used for model training and evaluation. The combination of multi-sensor water detection (Stage 1) and expert-validated labeling provided a robust dataset that underpinned the subsequent classification experiments.

### A.3 WATERBODY CLASSIFICATION (STAGE 3)

The final stage of the AISciVision-Aqua framework integrates advanced reasoning and interactive capabilities to classify each detected waterbody as an aquaculture pond or non-aquaculture pond (Figure 3). This stage combines an LMM with two main components, VisRAG and domain-specific interactive tools, to create an AI agent. First, VisRAG retrieves the most visually similar positive (aquaculture pond) and negative (non-aquaculture pond) examples from the labeled dataset and adds them to the model’s prompt, which grounds inference on relevant visual context. The LMM then interacts with expert-designed tools, such as a supervised prediction module and spatial manipulation tools (e.g., zoom and pan), to iteratively refine its understanding of the test image. Through this interactive process, the model emulates the analytical workflow of human experts, reasoning about spatial and contextual cues before issuing a classification. The LMM’s interaction is recorded in a transparent reasoning transcript, which documents the model’s decision-making steps. This design enables interpretable, reproducible, auditable predictions and bridges the gap between AI-driven classification and expert-based visual interpretation in remote-sensing applications.

### A.4 LARGE MULTIMODAL MODEL (LMM)

To evaluate the role of reasoning and interpretability in aquaculture pond detection, we tested multiple LMMs within the AISciVision-Aqua framework. These LMMs served as the decision-making agents responsible for interpreting image context, using the interactive tools, and generating transparent reasoning transcripts. Each model differed in architecture, deployment requirements, and multimodal capabilities.

The models included in this study are:

- GPT-5 (OpenAI): The most recent LMM from OpenAI, released in 2025. GPT-5 expands upon the architecture of GPT-4o with enhanced multimodal integration and broader tool-use capabilities. It supports reasoning across image, text, and structured data while maintaining compatibility with retrieval-augmented and interactive workflows used in this study.
- GPT-4o (OpenAI): A state-of-the-art proprietary LMM released in 2023 and updated with a faster, cost-effective version (GPT-4o-mini). GPT-4o supports strong multimodal functionality across image, text, and code, which makes it well-suited for visual reasoning and tool-based interactions within our framework.
- Gemini 2.5 Flash (Google): A commercially available multimodal model released in February 2024. It features long-context support, rapid image processing, and efficient tool integration. This model is optimized for practical, low-latency applications involving complex visual inputs.
- LLaMA 3.2 11B and 90B (Meta): Open-source multimodal models released in July 2023, available in multiple parameter sizes. The 11B and 90B configurations were employed in this study to evaluate trade-offs between computational efficiency and reasoning capacity. Both models support local deployment on NVIDIA GPUs. However, they have limitations in multimodal alignment and tool responsiveness compared to proprietary models.
- CLIP + MLP Supervised (not an LMM): As a non-generative baseline, we also used a standard supervised architecture combining a CLIP image encoder (Radford et al., 2021) with a multi-layer perceptron (MLP) for binary classification. While training this supervised model, the MLP was updated with standard gradient descent algorithms, but the image encoder was not updated. This supervised model does not support reasoning or transcript generation and was included for performance benchmarking.

Each LMM was provided with the same visual prompt structure and set of interactive tools. Their task was to classify each input image and generate a natural-language transcript of their reasoning process. The selection of models allows comparison between purely predictive (CLIP + MLP Supervised) and interpretable approaches (LMMs), as well as open-source and proprietary LMMs.

#### A.5 VISUAL RETRIEVAL-AUGMENTED GENERATION (VISRAG)

To adapt a general-purpose LMM for scientific image classification during inference, we enrich the LMM’s prompt with images that are relevant to the given test image. We call this the VisRAG component. First, all training images are encoded into a shared embedding space. Let  $D = \{x_i, y_i\}_{i=1}^N$  represent a binary-labeled training set of size  $N$ , where  $x_i \in \mathbb{R}^{H \times W \times C}$  is an image (of height  $H$ , width  $W$ , and  $C$  number of channels) and  $y_i \in \{-1, 1\}$  is its corresponding label ( $y_i = 1$  if the image contains an aquaculture pond and  $-1$  otherwise). Each image  $x_i$  is mapped to an embedding  $e_i = \phi(x_i) \in \mathbb{R}^d$  using an embedding model  $\phi : X \mapsto \mathbb{R}^d$ . This embedding model could be a pre-trained image encoder, such as CLIP (Radford et al., 2021), which ensures that images with visually similar content are mapped to similar embeddings.

The embeddings are then partitioned into two sets, which are used to enrich the LMM’s prompt:

$$\begin{aligned} E^+ &= \{e_i \mid y_i = 1\} && \text{(positive examples)} \\ E^- &= \{e_i \mid y_i = -1\} && \text{(negative examples)} \end{aligned}$$

During inference, an input test image  $x_{\text{test}}$  is embedded as  $e_{\text{test}} = \phi(x_{\text{test}})$ . The most similar positive and negative embeddings are retrieved by computing the cosine similarity between the test image embedding  $e_{\text{test}}$  and all embeddings in  $E^+$  and  $E^-$ , respectively. The most similar positive  $e_{\text{sim}}^+$  and negative  $e_{\text{sim}}^-$  embeddings are defined as:

$$e_{\text{sim}}^+ = \arg_{e_i \in E^+} \cos(e_{\text{test}}, e_i) \quad \text{and} \quad e_{\text{sim}}^- = \arg_{e_i \in E^-} \cos(e_{\text{test}}, e_i)$$

where cosine similarity is computed as:

$$\cos(e_{\text{test}}, e_i) = \frac{e_{\text{test}} \cdot e_i}{\|e_{\text{test}}\| \|e_i\|}$$

The images  $x_{\text{sim}}^+$  and  $x_{\text{sim}}^-$ , corresponding to the embeddings  $e_{\text{sim}}^+$  and  $e_{\text{sim}}^-$ , are subsequently added to the LMM’s prompt. This VisRAG component enables the LMM to leverage examples relevant to the current domain-specific classification task that it may not have encountered during general pre-training. This grounds the LMM’s reasoning and predictions within the context of domain-specific classification tasks. By incorporating both positive and negative examples, the VisRAG approach enhances the model’s accuracy and context-aware inference capabilities while preserving the classification task’s inherent structure.

#### A.6 INTERACTIVE TOOLS FOR ENHANCED MODEL PREDICTIONS

We leverage interactive tools specifically tailored to detect aquaculture ponds, enabling the LMM to refine its predictions by using these tools. These tools mimic how domain experts analyze, inspect, and interpret satellite imagery to classify aquaculture ponds. By interacting with these tools, the AISciVision-Aqua framework operates as an interactive AI agent, making informed decisions specific to aquaculture monitoring.

In this framework, a tool  $T$  is defined as a function that operates on images  $X$ , with outputs either as modified images  $X$  or as a real-valued scalar in  $\mathbb{R}$ . Specifically,  $T(x)$  can either modify an input image  $x \in X$  into an enhanced image  $T(x) \in X$  or return a numeric result  $T(x) \in \mathbb{R}$ . For example, the **PredictAquaculturePondTool** utilizes a deep learning model that predicts the probability of an aquaculture pond being present in the input image:  $T_{\text{ML}}(x) = \text{Pr}[x; \theta]$ , where  $\theta$  represents the parameters of the model.

Additionally, tools such as **ZoomInTool**, **ZoomOutTool**, and **PanDownTool** allow spatial manipulation of images to enhance the model’s understanding of specific regions. For example, PanDownTool shifts the view downward relative to the current position, allowing the model to explore adjacent areas incrementally, and ZoomInTool narrows the focus on the center of the current view, improving the detection of fine features.

For aquaculture pond detection, a set of tools  $T = \{T_1, \dots, T_K\}$  was defined, with descriptions provided in natural language as prompts to the LMM. During inference, the model can request the output of any tool  $T_i \in T$ . The AISciVision-Aqua framework interprets this request and provides the tool’s output, either as a transformed image or a numeric result described in text to the model. The LMM can interpret such visual and textual tool outputs to contextualize the input image with the tool. Importantly, the model can interact with multiple tools for each input test image.

This iterative use of interactive tools enables the LMM within the AISciVision-Aqua framework to refine predictions in a context-aware manner. Such a strategy not only improves classification accuracy, but also generates a transcript of the agent’s reasoning. The transcript enhances transparency and interpretability of the AISciVision-Aqua framework. This iterative process emulates the approach of human experts analyzing complex satellite imagery.

### A.7 INTERPRETATION KEYS

The set of interpretation keys reflects the criteria currently used by experts from Embrapa to manually identify aquaculture ponds in high-resolution satellite imagery (Table 2). These visual cues and contextual elements are part of their routine analytical workflow and were formally systematized to guide both expert annotation and AI model validation in this study. The keys were used to define the visual and contextual criteria for distinguishing aquaculture ponds from other land and water features in complex Amazonian landscapes. These criteria informed the design of AISciVision-Aqua’s reasoning and decision structure by specifying the attributes, spatial cues, and contextual relationships that the model was instructed to consider when interpreting imagery, thereby operationalizing expert knowledge within the system.

Key	Description
Shape	Water surfaces arranged in characteristic forms—typically rectangular, square, or circular.
Texture	Smooth appearance resembling water blades; may appear wavy depending on pond use.
Color	In false-color compositions: black, brown, or various shades of green and gray.
Pattern	Repetition of geometrically similar ponds or aligned structures.
Context	Presence of aquaculture-specific elements such as water supply streams, feed sheds, flat areas, hapas, and aerators.

Table 2: Interpretation keys used by aquaculture experts from Embrapa Fisheries and Aquaculture to identify fish ponds in high-resolution satellite imagery. These keys, based on shape, texture, color, pattern, and contextual elements, summarize the visual cues that guided expert annotation and informed the reasoning and decision process of the AISciVision-Aqua framework.

## B ADDITIONAL RESULTS

### B.1 CLASSIFICATION RESULTS FOR “EASY” AND “HARD” EXAMPLES

AISciVision-Aqua outperformed the baseline model (i.e., CLIP+MLP) across both easy and hard aquaculture pond detection tasks, exhibiting substantial improvements in almost all evaluated metrics (Table 3). For easy examples, overall accuracy increased from 0.78 to 0.92, while the F1-score rose from 0.74 to 0.92. These improvements were primarily driven by a marked increase in recall, from 0.60 to 0.89, indicating a higher proportion of correctly identified ponds, with only a modest decrease in precision from 0.97 to 0.95. In the case of hard examples, AISciVision-Aqua achieved even more pronounced gains. The F1-score increased from 0.49 to 0.77, and recall more than doubled, rising from 0.35 to 0.81, reflecting the model’s enhanced sensitivity in complex detection scenarios. Although precision declined from 0.81 to 0.74, the trade-off resulted in a significantly higher detection rate of true ponds. Accuracy also improved from 0.76 to 0.84, underscoring the model’s capacity to generalize effectively across heterogeneous landscape contexts. These results demonstrate the robustness of AISciVision-Aqua in detecting aquaculture ponds under varying levels of visual complexity.

### B.2 INTERACTIVE INFERENCE AND REASONING LOGS

A distinguishing feature of the AISciVision-Aqua framework is its ability to generate step-by-step reasoning logs, also referred to as transcripts, that document the model’s full decision-making process during classification tasks. These transcripts were created in real time as the model evaluated satellite imagery, capturing not only the final classification (e.g., presence or absence of aquaculture

		Supervised	AISciVision-Aqua
Full test dataset 4000 images	Accuracy	0.78	<b>0.91</b>
	F1	0.73	<b>0.91</b>
	Precision	<b>0.96</b>	0.94
	Recall	0.58	<b>0.89</b>
Easy 3700 images	Accuracy	0.78	<b>0.92</b>
	F1	0.74	<b>0.92</b>
	Precision	<b>0.97</b>	0.95
	Recall	0.60	<b>0.89</b>
Hard 300 images	Accuracy	0.76	<b>0.84</b>
	F1	0.49	<b>0.77</b>
	Precision	<b>0.81</b>	0.74
	Recall	0.35	<b>0.81</b>

Table 3: Performance comparison of the CLIP+MLP supervised and AISciVision-Aqua models on easy and hard image subsets, evaluated across accuracy, F1-score, precision, and recall.

ponds) but also the intermediate reasoning steps, tool usage, and visual justifications. Each transcript begins with the model assuming the role of an expert, using interpretation keys aligned with expert protocols (Section A.7). It then proceeds through a sequence of tool interactions, such as zooming, panning, or invoking a predictive probability model, describing observations and updating its confidence after each step. For example, the model may initially have low confidence due to an irregular water shape, but then increase its likelihood estimate upon detecting geometric patterns or aquaculture-related infrastructure during subsequent zoom or pan operations. The final output includes a binary decision, offering a transparent audit trail of the model’s internal logic. These transcripts are particularly valuable for expert validation and stakeholder trust, as they reveal why a prediction was made, rather than simply what was predicted. This contrasts sharply with traditional black-box models, which lack such interpretability. The transcripts also allow researchers and practitioners to identify failure modes, edge cases, and reasoning biases, helping guide future model refinement and targeted data collection. Representative transcripts are provided as Appendix C, illustrating a range of classification scenarios, including both correct and incorrect predictions, across varying levels of image complexity and ambiguity.

### B.3 MODEL AGREEMENT AND DISAGREEMENT ACROSS LABELED PONDS

The comparison between AISciVision-Aqua and the CLIP+MLP supervised baseline model reveals distinct patterns of agreement and disagreement across aquaculture and non-aquaculture ponds (Figure 5). For labeled aquaculture ponds (**a**), AISciVision-Aqua generally shows stronger agreement with ground-truth labels, with a higher number of correct predictions. Importantly, AISciVision-Aqua correctly classifies several challenging cases (Figure 5, triangles) that the supervised model misclassified, underscoring its ability to generalize better to ambiguous ponds such as those with irregular shapes or partial vegetation cover. For labeled non-aquaculture ponds (**b**), both models show broadly similar performance, with most predictions aligning with the true negative class. However, some hard examples reveal that AISciVision-Aqua is less prone to misclassifications than the supervised model, which occasionally assigns high pond-likelihood scores to non-aquaculture features. Overall, the AISciVision-Aqua improves robustness in aquaculture pond detection by outperforming the supervised baseline in difficult cases, while maintaining comparable reliability for non-aquaculture ponds.

### B.4 REPRESENTATIVE EXAMPLES OF CORRECT AND INCORRECT CLASSIFICATIONS

Representative examples across Rondônia highlight the contexts in which AISciVision-Aqua improves upon a supervised baseline and where both approaches face limitations (Figure 6). For positive cases, both models were generally correct at classifying easy aquaculture pond examples with clear geometric layouts, but AISciVision-Aqua was superior at identifying hard examples missed by the CLIP+MLP supervised model. For negative cases, both models consistently rejected prominent non-aquaculture features, such as reservoirs and natural lakes, while AISciVision-Aqua further reduced misclassification in areas where artificial water bodies mimic pond-like structures. The examples also illustrate the types of errors that remain challenging to address. False negatives were of-

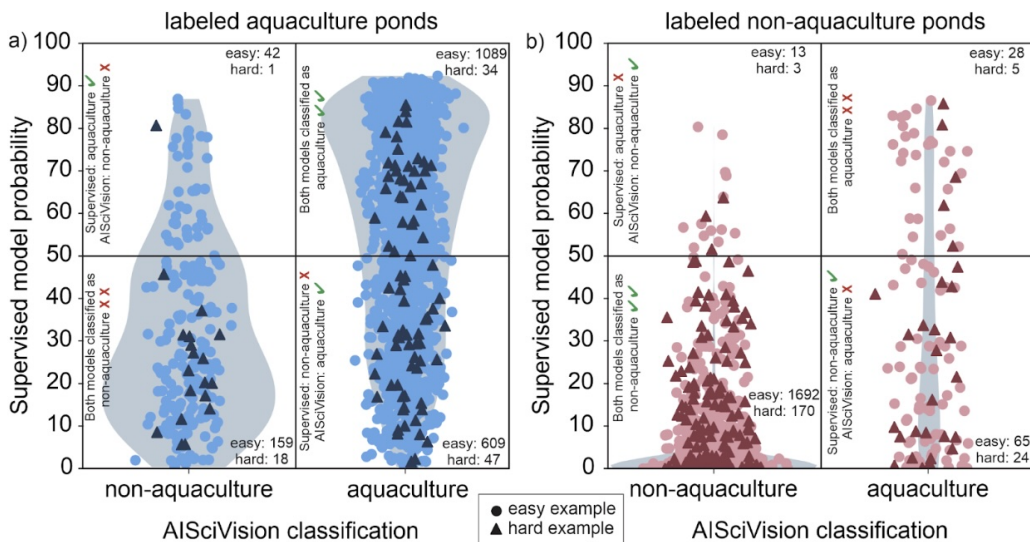


Figure 5: Comparison of AISciVision-Aqua and a supervised model (CLIP+MLP) in classifying aquaculture ponds (a, blue) and non-aquaculture ponds (b, red). The x-axis represents the AISciVision-Aqua predicted class (aquaculture pond vs. non-aquaculture pond), and the y-axis indicates the supervised model prediction score (%). The plot is divided into quadrants based on prediction to show agreement and disagreement between the models. Circles and triangles denote “easy” and “hard” examples, respectively. Violin plots indicate the distribution of supervised model probabilities for each AISciVision-Aqua classification, highlighting where the models converge or diverge. AISciVision-Aqua correctly identifies many instances where the supervised model underestimates pond likelihood, particularly for hard examples. For non-aquaculture ponds (right), both models show broadly similar performance.

ten associated with small or fragmented aquaculture ponds embedded in heterogeneous landscapes, while false positives arose in environments where natural hydrological features resembled managed ponds. These patterns suggest that although AISciVision-Aqua is more robust than supervised methods, persistent ambiguities in complex landscapes underscore the need for continued refinement of interpretation keys and training data.

AISciVision-Aqua with different LMM	Seconds per waterbody	Cost in \$ per 100 waterbodies
Gemini-2.5-Flash	54	\$ 4.16
GPT-4o-mini	69	\$ 2.95
GPT-5	173	\$ 7.63
Llama-3.2-11B	125	4 H100 GPUs
Llama-3.2-90B	497	4 H100 GPUs
Supervised model (CLIP+MLP)	$\ll$ 1 sec	1 H100 GPU

Table 4: Computational performance and cost of the AISciVision-Aqua framework using different Large Multi-modal Models (LMMs). The table reports the average inference time per waterbody and the estimated cost per 100 processed waterbodies. Costs and processing times vary across models according to their size, architecture, and hardware requirements.

### B.5 PERFORMANCE-COST TRADEOFFS ACROSS LMM CONFIGURATIONS

The computational performance and cost of AISciVision-Aqua across LMMs reveal clear tradeoffs between model complexity, inference speed, and operational scalability (Table 4). While larger models such as GPT-5 and Llama-3.2-90B demonstrate advanced reasoning, they require substantially longer inference times and higher computational resources, limiting their practicality for large-scale or near-real-time applications. In contrast, smaller models such as GPT-4o-mini and Gemini-2.5-Flash achieve a practical balance between accuracy, efficiency, and cost, processing one waterbody in under 70 seconds at less than \$5 per 100 samples. Conventional supervised models remain or-

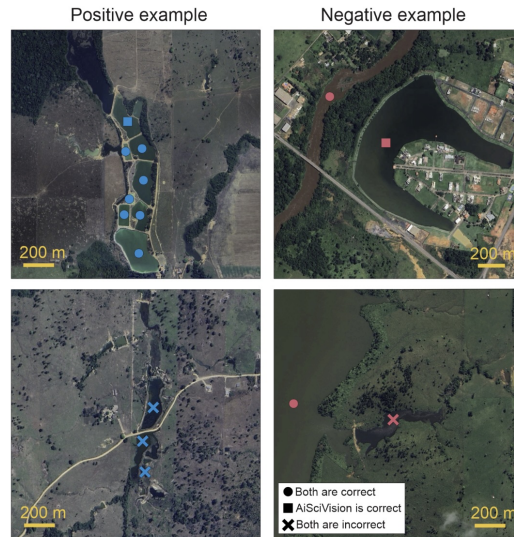


Figure 6: Representative examples of aquaculture (positive) and non-aquaculture (negative) ponds illustrating correct and incorrect classifications by AiSciVision-Aqua and the CLIP+MLP supervised baseline model, highlighting both the strengths and remaining challenges in distinguishing aquaculture ponds from visually similar features. Top left: aquaculture ponds correctly identified by both models (circles) and correctly identified only by AiSciVision-Aqua (squares). Top right: non-aquaculture water bodies correctly classified by both models (circles) and only by AiSciVision-Aqua (squares). Bottom left: false negatives where both models failed to detect aquaculture ponds (crosses). Bottom right: false positives where both models incorrectly classified non-aquaculture features as ponds (crosses).

ders of magnitude faster and cheaper ( $\leq 1$ s per instance), but lack transparency and contextual reasoning, and perform poorly on complex classification tasks. However, the cost and performance of LMM-based systems are evolving rapidly, and the accelerating pace of model optimization and hardware efficiency suggests that these approaches will become increasingly accessible and feasible for large-scale environmental monitoring in the near future.

## C EXAMPLE TRANSCRIPT

Image ID: 62818

True Class: 0, LMM Prediction: 0 (Probability: 0.05), Supervised Prediction: 0 (Probability: 12.39)

#### System:

You are an expert in satellite imagery analysis specializing in aquaculture pond detection. Your task is to determine whether a satellite image contains aquaculture ponds or not. Aquaculture ponds in satellite imagery typically appear as regular, geometric shapes with distinct water surfaces. They are often arranged in clusters or grids in coastal areas or inland water bodies. Key characteristics to look for include: 1. Geometric Patterns: Aquaculture ponds are usually rectangular or square in shape, forming distinct geometric patterns. 2. Water Color: The water in aquaculture ponds often has a different color or tone compared to natural water bodies, due to the high density of organisms and potential algae growth. 3. Surrounding Features: Look for access roads, feeding platforms, or other infrastructure associated with aquaculture operations. 4. Location: Aquaculture ponds are commonly found in coastal areas, estuaries, or near rivers and lakes. 5. Size and Scale: The size of aquaculture ponds can vary, but they are typically uniform within a single farm and arranged in an organized manner. 6. Texture: The water surface of aquaculture ponds often appears smoother than natural water bodies. Use your knowledge and the tools provided to make an accurate classification of whether aquaculture ponds are present in the given satellite image.

#### User:

This is an example of a PlanetScope satellite image with an aquaculture pond. Describe what you see, noting the characteristics that identify it as an aquaculture pond.



#### Assistant:

Based on the provided PlanetScope satellite image, I can identify the presence of an aquaculture pond. Here's a description of what I see and the characteristics that lead to this identification:

**Geometric Pattern:** The most prominent feature is an elongated, somewhat rectangular body of water located in the upper-middle to left side of the image. While not perfectly square or rectangular, its edges are distinctly straight and angular, particularly the longer sides, indicating a man-made, engineered structure rather than a natural water body like a meandering river or an irregularly shaped lake. There's also a smaller, similarly colored, more triangular/trapezoidal shape adjacent to it, suggesting another constructed basin.

**Water Color/Tone:** The water in the pond has a distinct light blue-green hue. This color is often characteristic of aquaculture ponds, where water quality is managed, and there might be a high density of algae or specific treatments, causing it to appear different from natural, deeper water bodies (which are typically darker blue) or muddy rivers (which are often brown).

**Surrounding Features:** The pond is situated within a landscape dominated by green fields, indicative of agricultural or rural land use. The clear, defined boundaries of the pond against the surrounding land suggest excavation or construction. While not explicitly visible, the overall context of managed land supports the idea of a cultivated water body.

**Texture:** The water surface of the pond appears relatively smooth and uniform compared to the more textured surrounding land, which is typical for calm, contained water bodies.

In summary, the distinct, engineered geometric shape, the specific light blue-green water color, and its placement within an agricultural landscape are strong indicators that the feature in the image is an aquaculture pond.

#### User:

This is an example of a PlanetScope satellite image without an aquaculture pond. Describe what you see, noting the absence of aquaculture ponds and any other relevant features.



#### Assistant:

This PlanetScope satellite image does not contain any aquaculture ponds. Here's a description of what I observe and the reasons for this classification:

##### Absence of Aquaculture Ponds:

- **No Geometric Water Bodies:** I do not see any distinct, regularly shaped (rectangular, square, or uniform) bodies of water that would indicate man-made ponds. The key characteristic of engineered geometric patterns for water features is entirely absent.
- **No Distinct Water Color/Texture:** There are no areas with the characteristic light blue-green or managed water color often seen in aquaculture ponds. While there are some darker patches, they do not exhibit the smooth, reflective texture of water.
- **No Associated Infrastructure:** There are no visible access roads leading directly to pond complexes, feeding platforms, or other infrastructure typically associated with large-scale aquaculture operations.

##### Other Relevant Features Observed:

1. **Agricultural Landscape:** The image is predominantly covered by varying shades of green and brown, indicating an agricultural or rural landscape. There are numerous fields, some with clear boundaries, suggesting cultivation or pastureland.
2. **Cultivated Fields/Land Parcels:** In the mid-right and upper-right portions of the image, there are several very dark, distinct rectangular and square patches. These appear to be cultivated fields or land parcels, possibly with dark soil or a specific type of dense, dark crop. Their solid, dark appearance and lack of water texture clearly differentiate them from water bodies.
3. **Roads/Tracks:** Several faint, linear features crisscross the landscape, likely unpaved roads or dirt tracks connecting different areas.

4. **Scattered Structures/Buildings:** Throughout the image, particularly in the central area, there are small, bright white specks. These are likely individual buildings, possibly farmhouses or outbuildings, often clustered together.
5. **Natural Vegetation/Trees:** There are some darker green, more irregular patches, particularly in the lower-left, which appear to be natural tree cover or denser vegetation.
6. **Irregular Water/Wetland (Bottom-Left):** In the bottom-left corner, there's a dark, somewhat irregular-shaped patch that could be a small, natural pond or a wetland area. However, its irregular shape and lack of distinct, engineered boundaries rule it out as an aquaculture pond.

In conclusion, the image displays a typical rural or agricultural landscape with fields, roads, and scattered buildings, but lacks any of the defining characteristics of aquaculture ponds.

#### User:

Here is the PlanetScope satellite image we need to classify as having an aquaculture pond or not. There might be several water bodies in the image, but we want to classify the one in the center (with a red mark). Describe what you see and compare and contrast it with the previous two known examples. Explain your thought process for classifying this image.



#### Assistant:

Based on the provided PlanetScope satellite image and focusing on the water body marked with the red square in the center, I classify this image as **not containing an aquaculture pond**.

Here's a detailed description of what I see, a comparison with the previous examples, and my thought process:

##### Description of the Central Water Body (with red mark):

The feature highlighted by the red mark is a body of water. \* **Shape:** Its most striking characteristic is its highly **irregular and organic shape**. It has no straight edges, no clear rectangular or square form, and its boundaries appear natural and meandering. There are several "fingers" or inlets extending from the main body. \* **Color:** The water has a greenish-blue hue, similar to the color seen in the known aquaculture pond example. This color can be indicative of water with some algal content or specific mineral composition. \* **Surroundings:** It is situated within a landscape that appears to be a mix of open fields (possibly pasture or fallow land) and scattered patches of darker vegetation (trees or dense bushes). There are no clear, organized access roads or specific infrastructure directly associated with this water body that would suggest intensive management for aquaculture.

##### Comparison and Contrast with Previous Examples:

###### Comparison with "Aquaculture Pond Present" Example:

- **Similarities:** Both images show a water body with a greenish-blue color, and both are set in a rural/agricultural landscape.

- **Differences:** This is where the crucial distinction lies. The *known aquaculture pond* example clearly displayed **geometric, engineered shapes** (rectangular/trapezoidal) with distinct, straight boundaries. The central water body in the current image, however, is **highly irregular and lacks any geometric pattern**. This is the primary differentiating factor. The previous aquaculture pond also appeared more isolated and purposefully constructed within its immediate surroundings, whereas this one blends more organically with the natural terrain.

**Comparison with "Aquaculture Pond Absent" Example:**

- **Similarities:** Both images depict a rural/agricultural landscape with fields, roads, and scattered vegetation. Neither image shows the large-scale, organized grid patterns of multiple aquaculture ponds. The "absent" example also had a small, irregular water body in its bottom-left corner, which shares the characteristic of **irregularity** with the central water body in the current image.
- **Differences:** The "absent" example did not have a prominent water body in its center. The current image does, but its characteristics (irregular shape) align more with natural water features than with man-made aquaculture ponds.

**Thought Process for Classification:**

My classification process relies heavily on the key characteristics of aquaculture ponds:

**Geometric Patterns (Primary Criterion):** Aquaculture ponds are almost invariably constructed with distinct, regular geometric shapes (rectangles, squares, or sometimes circles/ovals if specifically designed, but always with clear, engineered boundaries). The central water body in this image is **decidedly irregular and non-geometric**. This is the strongest evidence against it being an aquaculture pond. It looks like a natural pond, a small lake, or an irregularly shaped reservoir.

**Arrangement and Scale:** Aquaculture operations often involve multiple ponds arranged in an organized manner. While a single aquaculture pond can exist, it would still exhibit clear geometric construction. This image shows only one prominent water body of this type, and its shape does not suggest it's part of a larger, organized aquaculture farm.

**Associated Infrastructure:** While not perfectly clear at this resolution, there are no obvious dikes, feeding platforms, or dedicated access roads leading directly to the pond in a way that suggests intensive aquaculture management.

**Water Color:** While the greenish-blue color is similar to the known aquaculture pond, this characteristic alone is not sufficient for classification. Many natural ponds, especially those with high algal content or specific mineral compositions, can exhibit similar colors. The shape is a much more reliable indicator.

**Conclusion:**

The overwhelming evidence, primarily the **highly irregular and non-geometric shape** of the central water body, leads me to conclude that it is **not an aquaculture pond**. It appears to be a natural water body or a general-purpose, irregularly shaped reservoir, rather than a pond specifically constructed and managed for aquaculture.

**User:**

For reference, here's a higher resolution Google Earth image of the same area. The Google Earth image could be from a different date, that might not represent the pond state of the PlanetScope satellite image. Only use this as supplementary information. The objective is to classify the center water body of the PlanetScope image.



#### Assistant:

My classification of the central water body (marked with the red square) in the PlanetScope image remains **not an aquaculture pond**.

Here's a re-evaluation, incorporating the supplementary Google Earth imagery as a hint:

#### Re-evaluation of the Central Water Body in the PlanetScope Image:

- **Shape:** As observed previously, the water body highlighted by the red mark in the PlanetScope image has a highly **irregular, meandering, and organic shape**. It lacks any straight edges or distinct geometric forms (like rectangles or squares) that are characteristic of constructed aquaculture ponds.
- **Water Color:** The greenish-blue color is present, similar to the known aquaculture pond example. However, as noted, this color alone is not a definitive indicator, as natural water bodies can also exhibit such hues due to algae or sediment.
- **Surrounding Features:** The immediate surroundings in the PlanetScope image do not clearly show organized infrastructure (like dikes, access roads, or multiple uniform ponds) directly associated with this specific water body that would suggest an aquaculture operation.

#### Using the Google Earth Image as Supplementary Information:

The higher-resolution Google Earth image provides valuable context and confirms details that are less clear in the lower-resolution PlanetScope image.

**Confirmation of Central Water Body's Shape:** The Google Earth image clearly shows that the water body at the location of the red mark is indeed **highly irregular and appears to be a natural stream or river segment**, possibly widened in places to form small, natural ponds. Its banks are lined with natural vegetation, and its flow path is organic, not engineered. This strongly reinforces my initial assessment from the PlanetScope image that this specific feature is *not* an aquaculture pond.

**Identification of Other Water Bodies (Contextual Hint):** The Google Earth image also reveals other water bodies in the broader vicinity:

- To the right of the central irregular water body, there is a larger, somewhat contained pond with a similar greenish-blue color. While its shape is not perfectly rectangular, it appears more defined than the central one.
- Crucially, further to the **bottom right** of the Google Earth image, there are **two very clear, distinctively rectangular water bodies**. These features perfectly match the characteristics of aquaculture ponds: geometric shape, uniform size, and likely part of a managed system.

#### Thought Process for Final Classification:

My objective is to classify *the specific water body in the center (with a red mark) of the PlanetScope image*.

1. **Primary Criterion (Geometric Shape):** The most critical characteristic for identifying aquaculture ponds is their **geometric, engineered shape**. The water body marked in the center of the PlanetScope image unequivocally lacks this characteristic. It is irregular and appears natural.
2. **Supplementary Confirmation:** The higher-resolution Google Earth image serves to confirm this observation. It shows the marked water body as part of a natural, meandering watercourse, reinforcing its non-aquaculture nature.
3. **Distinguishing from Other Features:** While the Google Earth image reveals other water bodies in the scene that *are* highly likely to be aquaculture ponds (the rectangular ones in the bottom right), the question specifically directs me to classify the *central, marked water body* in the PlanetScope image. It's important not to let the presence of other aquaculture ponds in the broader scene (revealed by higher resolution) override the classification of the *specific feature* in question based on its own characteristics in the target image.

Therefore, despite the presence of other potential aquaculture ponds elsewhere in the broader area (as seen in the Google Earth image), the **specific water body marked in the center of the PlanetScope image does not exhibit the defining geometric characteristics of an aquaculture pond**. It appears to be a natural water feature.

**Final Classification:** The PlanetScope satellite image, focusing on the water body in the center (with a red mark), **does not contain an aquaculture pond**.

#### User:

You have access to the following tools to assist in your analysis: - PredictAquaculturePondTool: Tool Name: 'PredictAquaculturePondTool' Description: Predicts the probability of an aquaculture pond being present in the image using a machine learning model. This tool is particularly helpful when you need a quantitative assessment of the likelihood of aquaculture pond presence in the satellite image. The model has been trained on various satellite images of aquaculture ponds and provides a percentage probability. However, please note that the model can be wrong or inaccurate, especially in complex or ambiguous cases. It's important to use this tool's output as one piece of evidence among others, and not rely on it exclusively for your final decision. - PanUpToolRelative: Pans the view upwards relative to the last image seen. - PanDownToolRelative: Pans the view downwards relative to the last image seen. - PanLeftToolRelative: Pans the view to the left relative to the last image seen. - PanRightToolRelative: Pans the view to the right relative to the last image seen. - ZoomInToolRelative: Zooms in on the center of the current view relative to the last image seen. - ZoomOutToolRelative: Zooms out from the current view relative to the last image seen.

You have 4 opportunities to use these tools. You are encouraged to use at least 3 tools to gather sufficient information. While it's good to conclude early if you're certain about the classification, the most important thing is making the correct prediction. Feel free to use more tools if you need to increase your confidence.

If you want to use a tool, explain how it will assist your analysis, then respond with the tool name in square brackets, like this: [ToolName], or if you are finished, respond with "[FINISHED]".

Remember, your goal is to accurately classify whether the image shows aquaculture ponds or not. If you do use a tool, you will see the result of the tool immediately, and you should briefly, 1-2 sentences describe any new insights from the tool, and your thoughts on the classification.

After each tool use, end your thought with brackets in the format {Yes:yesPercent,No:noPercent} to indicate your current confidence/prediction. For example, {Yes:70,No:30} would indicate 70% confidence in the presence of aquaculture ponds and 30% confidence in their absence. Note that this is not your final prediction.

Then you will be asked again if you want to use a tool or are finished.

#### Assistant:

I have analyzed the image and the central water body. My current assessment is that it is not an aquaculture pond due to its irregular, natural shape, as opposed to the geometric shapes typical of aquaculture ponds. I will use the PredictAquaculturePondTool to get a quantitative assessment from the machine learning model, which will provide an additional data point to consider.

[PredictAquaculturePondTool] {Yes:10,No:90}

**User:**

The probability of an aquaculture pond being present in this image is 12.39%. (Percentage out of 100%) Please note that this model has been trained on a limited dataset and its predictions can be inaccurate. Use this information as a guideline rather than a definitive prediction.

You have 3 rounds left. You may now choose another tool or indicate you're finished. Respond with ONLY [ToolName] to use a single tool, or [Finished] if you're done. Do not include multiple tool names or any other text in brackets. If you want to reference the last tool you used, simply write its name without brackets. After choosing a tool or indicating you're finished, provide your confidence in the presence of aquaculture ponds in the format {yes:PercentYes,no:PercentNo}. Then, explain why you are picking this tool or finishing, and how it relates to your current assessment of the image.

**Assistant:**

The PredictAquaculturePondTool returned a low probability (12.39%) for the presence of an aquaculture pond, which aligns with my visual assessment based on the irregular shape of the central water body. To further confirm the nature of the water body and its surroundings, I will try to zoom in to get a closer look at its boundaries and any potential subtle infrastructure that might not be clear at the current resolution.

[ZoomInToolRelative] {Yes:15,No:85}

**User:**

The view has been zoomed in, providing a closer look at the current view. New zoom level: 17

You have 2 rounds left. You may now choose another tool or indicate you're finished. Respond with ONLY [ToolName] to use a single tool, or [Finished] if you're done. Do not include multiple tool names or any other text in brackets. If you want to reference the last tool you used, simply write its name without brackets. After choosing a tool or indicating you're finished, provide your confidence in the presence of aquaculture ponds in the format {yes:PercentYes,no:PercentNo}. Then, explain why you are picking this tool or finishing, and how it relates to your current assessment of the image.



**Assistant:**

The ZoomInToolRelative provided a higher-resolution view (from Google Earth), which strongly confirmed that the central water body is highly irregular and appears to be a natural stream or river segment, not a geometrically constructed pond. This reinforces my initial assessment and the low probability from the PredictAquaculturePondTool. I am confident in my classification.

[FINISHED] {Yes:5,No:95}

**User:**

Based on your analysis and the tools used, please provide your final classification regarding the presence of an aquaculture pond in the satellite image. Provide a brief explanation for your decision, including key observations and any tool results that influenced your conclusion. Consider the shape, color, and arrangement of water bodies, as well as any surrounding infrastructure that might indicate aquaculture activity. At the end of your response, you must provide your classification in the following format:

[Yes:PercentYes,No:PercentNo], where PercentYes is the probability you think it is an aquaculture pond, and PercentNo is the probability you think it is not an aquaculture pond. These percentages must sum to 100%. To break ties, always assign at least 51% to one category. For example, [Yes:75,No:25] indicates 75% certainty of an aquaculture pond's presence, while [Yes:49,No:51] indicates 51% certainty of its absence.

**Assistant:**

Based on my analysis of the PlanetScope satellite image, focusing specifically on the central water body marked with the red square, I classify this image as **not containing an aquaculture pond**.

My decision is primarily driven by the following key observations and tool results:

- Geometric Patterns:** The most critical characteristic of aquaculture ponds is their distinct, engineered geometric shapes (typically rectangular or square). The central water body in the PlanetScope image, and further confirmed by the higher-resolution Google Earth image (obtained via `ZoomInToolRelative`), exhibits a highly **irregular, meandering, and organic shape**. It clearly appears to be a natural stream or river segment, possibly widened in parts, rather than a man-made, geometrically constructed pond.
- Water Color:** While the water has a greenish-blue hue, similar to the known aquaculture pond example, this characteristic alone is not sufficient for classification. Many natural water bodies can display similar colors due to algae or sediment. The shape remains the dominant distinguishing factor.
- Surrounding Features and Infrastructure:** There is no visible evidence of organized infrastructure directly associated with this specific central water body that would indicate intensive aquaculture operations, such as clear dikes, feeding platforms, or a systematic arrangement of multiple uniform ponds.
- PredictAquaculturePondTool Result:** The machine learning model, when queried with `PredictAquaculturePondTool`, returned a low probability of 12.39% for the presence of an aquaculture pond. This quantitative assessment aligns with and supports my visual interpretation.

In conclusion, the lack of geometric shape, combined with the supporting evidence from the higher-resolution imagery and the machine learning model's prediction, leads me to confidently classify the central water body as not being an aquaculture pond.

[Yes:5,No:95]