# Documentation experiments

This document described three experiments and their results: the Amounts and Likelihood Adjusted Quantity Analysis, the prediction of deforestation classes, and the analysis of isolated pixels. Unfortunately, none of these methods were effective enough to be suitable for implementation in the FF project.

## Amounts and Likelihood Adjusted Quantity Analysis

An XGBoost model (with objective = "reg:squarederror") was trained using data from January 1, 2022, to December 1, 2022, to predict the number of integrated alerts per 400 x 400 m pixel. This model was then used to forecast deforestation amounts for all FF countries for the period from June 1, 2023, to December 1, 2023. These predictions were compared against confidence predictions derived from a binary model trained on the same period.

Additionally, the Likelihood Adjusted Quantity (LAQ) was calculated by multiplying the confidence predictions with the amount predictions. LAQ effectively combines both the predicted amount of deforestation and its likelihood into a single metric.

To determine the best predictive variable, the three variables (LAQ, Amounts, and Confidence) were compared by evaluating their correlation with the ground truth. Moreover, the average F0.5 scores for each variable were calculated and compared.

A graph of blue and orange bars

Description automatically generated

Figure 1 The average correlation between LAQ, Amounts, Confidence and the Actual Amounts

Only small differences are observed between the correlations of the three variables and the ground truth of the number of integrated alerts per pixel (Figure 1). The LAQ shows the highest correlation, followed by the confidence.

A graph on a white background

Description automatically generated

A graph on a white background

Description automatically generated

Figure 2 The F05-score for deforestation prediction using a different Threshold for (A) the Amount prediction and (B) The Likelihood Adjusted Quantity (LAQ).

Previous research indicated that a confidence threshold of 0.5 is optimal for predicting deforestation based on the binary model. A small experiment was conducted to determine the best thresholds for binary predictions using the Amounts and LAQ metrics. The results showed that the optimal threshold is 6 for Amounts and 3 for LAQ (Figure 2). These thresholds were used in the subsequent analysis.

A blue and orange pie chart

Description automatically generated

Figure 3 The method resulting in the highest F05-scores considering all FF countries.

A graph of blue and orange vertical bars

Description automatically generated

Figure 4 The F05 scores for the three methods (LAQ, Amounts and Confidence) and for 5 of the FF focus countries.

Considering all FF countries, the Likelihood Adjusted Quantity (LAQ) was the most effective method for predicting the occurrence of deforestation (Figure 3). However, for the FF focus countries, the F0.5-score obtained from the Confidence predictions generally outperforms the F0.5-scores of both the LAQ and Amount predictions (Figure 4). Therefore, the use of both Amounts and LAQ might not be necessary.

## Predicting Deforestation Classes

### A model was developed to predict three classes of deforestation in Colombia using data from January 1, 2022, to December 1, 2022. Initially, only pixels identified by the binary model as showing deforestation were selected. These selected pixels were then categorized into different classes using thresholds determined by the 0.5 and 0.9 quantiles of the training labels. These thresholds were subsequently applied to categorize both the training and testing labels, allowing for the prediction and classification of three levels of deforestation.

Table 1 Confusion Matrix

|  |  | **Reference** | | |
| --- | --- | --- | --- | --- |
|  |  | **Class: 0** | **Class: 1** | **Class: 2** |
| **Prediction** | **Class: 0** | 84239 | 17611 | 3944 |
| **Class: 1** | 25584 | 17130 | 4700 |
| **Class: 2** | 2701 | 2898 | 2881 |

Table 2 Performance Metrics Summary

| **Metric** | **Value** |
| --- | --- |
| **Accuracy** | 0.6448 |
| **95% CI** | (0.6424, 0.6471) |
| **No Information Rate** | 0.6959 |
| **P-Value [Acc > NIR]** | 1 |
| **Kappa** | 0.2484 |
| **McNemar's Test P-Value** | < 2e-16 |

Table 3 Statistics by Class

| **Statistic** | **Class: 0** | **Class: 1** | **Class: 2** |
| --- | --- | --- | --- |
| **Sensitivity** | 0.7486 | 0.4551 | 0.24998 |
| **Specificity** | 0.5616 | 0.7559 | 0.96271 |
| **Pos Pred Value** | 0.7963 | 0.3613 | 0.33974 |
| **Neg Pred Value** | 0.4940 | 0.8205 | 0.94358 |
| **Precision** | 0.7963 | 0.3613 | 0.33974 |
| **Recall** | 0.7486 | 0.4551 | 0.24998 |
| **F1 Score** | 0.7717 | 0.4028 | 0.28803 |
| **Prevalence** | 0.6959 | 0.2328 | 0.07128 |
| **Detection Rate** | 0.5210 | 0.1059 | 0.01782 |
| **Detection Prevalence** | 0.6543 | 0.2932 | 0.05245 |
| **Balanced Accuracy** | 0.6551 | 0.6055 | 0.60635 |

The model's overall accuracy is 64.48% (Table 2), which is below the No Information Rate of 69.59% (Table 2), indicating that it performs worse than simply predicting the most frequent class. The Kappa statistic of 0.2484 suggests low agreement beyond chance, and the McNemar's test p-value indicates significant differences in prediction errors across classes. The model performs relatively well for Class 0 with a sensitivity of 74.86% but struggles with Class 1 and Class 2, showing low sensitivity of 45.51% and 24.998%, respectively (Table 3), indicating that it fails to correctly identify these classes.

## Isolated Pixel Analysis

An analysis was conducted on isolated pixels from the model’s deforestation predictions. These isolated pixels are instances where deforestation is predicted, but none of the surrounding pixels indicate deforestation. The hypothesis was that removing these isolated pixels might improve the overall F0.5 score, a metric that emphasizes precision more than recall. This hypothesis was tested across all tiles for the date January 1, 2023. The results showed that removing isolated pixels only when their confidence value was below a threshold of 0.55 yielded the best outcomes. At this threshold, 54% of the isolated pixels were identified as False Positives (FP), representing 3.07% of the total FPs. Precision improved slightly from 68.12% before removal to 68.53% after removal. However, recall decreased marginally from 46.13% to 45.57%. Despite the small drop in recall, the slight increase in precision resulted in an overall increase of the F0.5 score by 0.06%.