CLOUD COVER DETECTION IN MULTI SPECTRAL IMAGERY

Authors– Ram sai Reddy Atla, Shubham Dogra, Roberto Flacco, Serah Johnson, Ainu Joshy, Esha Trivedi

Introduction

This Report evaluates five alternative machine learning algorithms for cloud cover detection in multi spectral satellite images. Project objective is to determine the best machine learning algorithm according to various evaluation criteria subject to computational complexity and efficiency constraints. Two groups of dataset images are tested by the machine learning algorithms which include four trained images and four untrained images. The machine learning algorithms utilize a ten-fold repeated cross validation repeated three times to find optimal hyperparameter setting for performance maximisation.

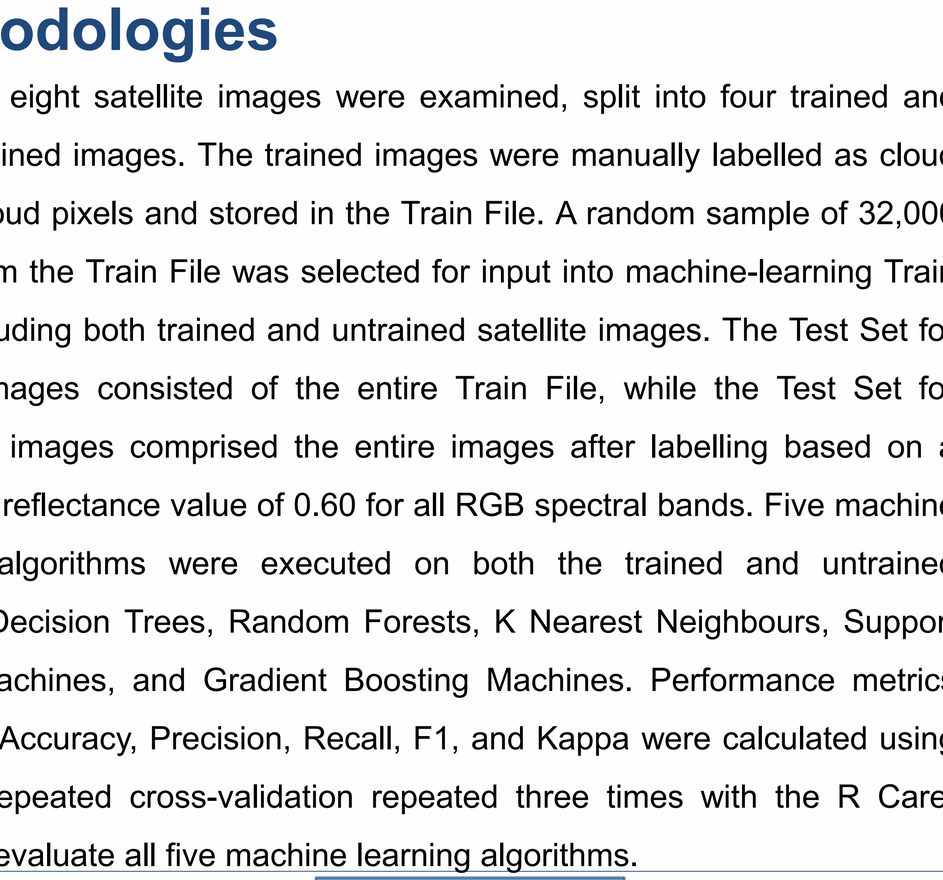
Aims

The goal of the Project is to deliver a machine-learning algorithm to SmartSat CRC for the detection of cloud cover in multi-spectral satellite images together with its associated documentation. Development of five competing machine learning algorithms for cloud cover detection are evaluated, from which the best algorithm is selected and taken forward to meet client business requirements.

Methodologies

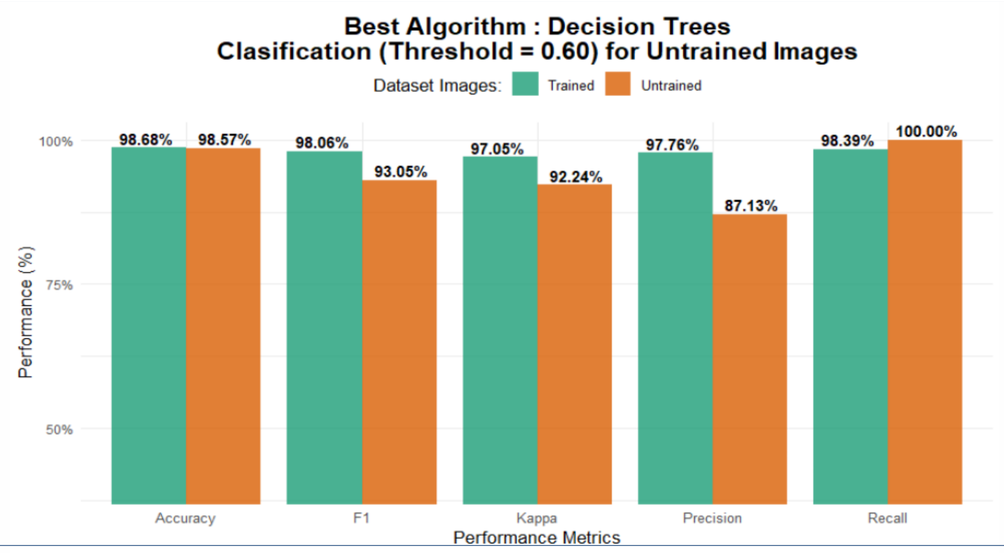
A total of eight satellite images were examined, split into four trained and four untrained images. The trained images were manually labelled as cloud or non-cloud pixels and stored in the Train File. A random sample of 32,000 pixels from the Train File was selected for input into machine-learning Train Sets, including both trained and untrained satellite images. The Test Set for trained images consisted of the entire Train File, while the Test Set for untrained images comprised the entire images after labelling based on a threshold reflectance value of 0.60 for all RGB spectral bands. Five machine learning algorithms were executed on both the trained and untrained images: Decision Trees, Random Forests, K Nearest Neighbours, Support Vector Machines, and Gradient Boosting Machines. Performance metrics including Accuracy, Precision, Recall, F1, and Kappa were calculated using ten-fold repeated cross-validation repeated three times with the R Caret library to evaluate all five machine learning algorithms.

FLOW CHART OF MODEL IMPLEMENTATION



Results

Five machine learning algorithms were evaluated for their performance on trained and untrained images. Models were rated on for best all-round performance according to the following criteria: Performance Consistency Across Datasets: The ability of a model to perform well not just on trained data but also on untrained data which indicates good generalization where results can vary significantly. High Scores Across Metrics: A model consistently scoring high across various metrics (Accuracy, Precision, Recall, F1, and Kappa) is likely reliable and versatile. Minimal Performance Drop from Trained to Untrained: A smaller drop in performance metrics from trained to untrained datasets suggests a model is less likely to overfit and more likely to handle new dataset images more effectively. Diagram below relates to the best rated machine learning algorithm where threshold reflectance value was set to 0.60 for untrained images. Decision Trees demonstrated the highest level of consistency between trained and untrained datasets across all performance metrics. The drop in performance was relatively minor compared to other algorithms, which indicates better generalisation. Precision showed lowest performance of 87.13% for untrained images while the highest was 100.00% for Recall for untrained dataset images as shown in diagram results below.

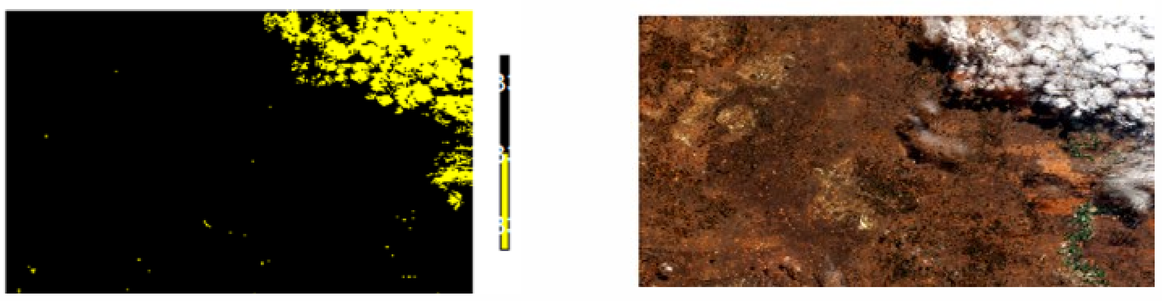


Conclusion

Five competing machine learning algorithms were examined to find the best model for detecting cloud cover in satellite imagery. The Decision Trees algorithm emerged as the best model for this task, based on comprehensive analysis and performance metrics evaluation. This conclusion was derived from an extensive assessment of various criteria, including accuracy, precision, recall, F1 score, and Kappa metrics.

Predicted Cloud - WAIKERIE
[Yellow - Cloud];
[Black - Non-Cloud].

Actual Cloud - WAIKERIE
RGB (Red, Green, Blue)



Comparison of Predicted vs. Actual Cloud Cover for Untrained Images