DS 6030– Disaster Relief

Group 1:

Maneesh Malpeddi, Pranav Sridhar, Nakisha Fouch, Sae’von Palmer

Introduction

Goal:

- Locate displaced persons in post-earthquake Haiti by identifying makeshift shelters (blue tarps) in aerial imagery, to deliver critical aid in time.
- Develop a classification model that can accurately and efficiently detect blue tarp pixels in a highly imbalanced dataset, using RGB values from aerial imagery.

Challenge:

- Severe class imbalance: Blue tarp pixels ~ 3.2% of data
- False negatives have a high cost – where missing tarps means missing aid

Strategy:

- Standardize RGB labeling across datasets
- Use stratified sampling and cross-validation
- Evaluate 5 models: Logistic Regression, LDA, QDA, KNN, SVM
- Focus on metrics that reflect class imbalance: Recall, F1, ROC AUC

Data Overview

Source

- Aerial imagery over Haiti post-2010 earthquake

Data

- Pixel-level color values + class label (Blue Tarp / Not Blue Tarp)

Challenge

- Color channels labeled as B1, B2, B3 — not standard RGB

Initial Approach

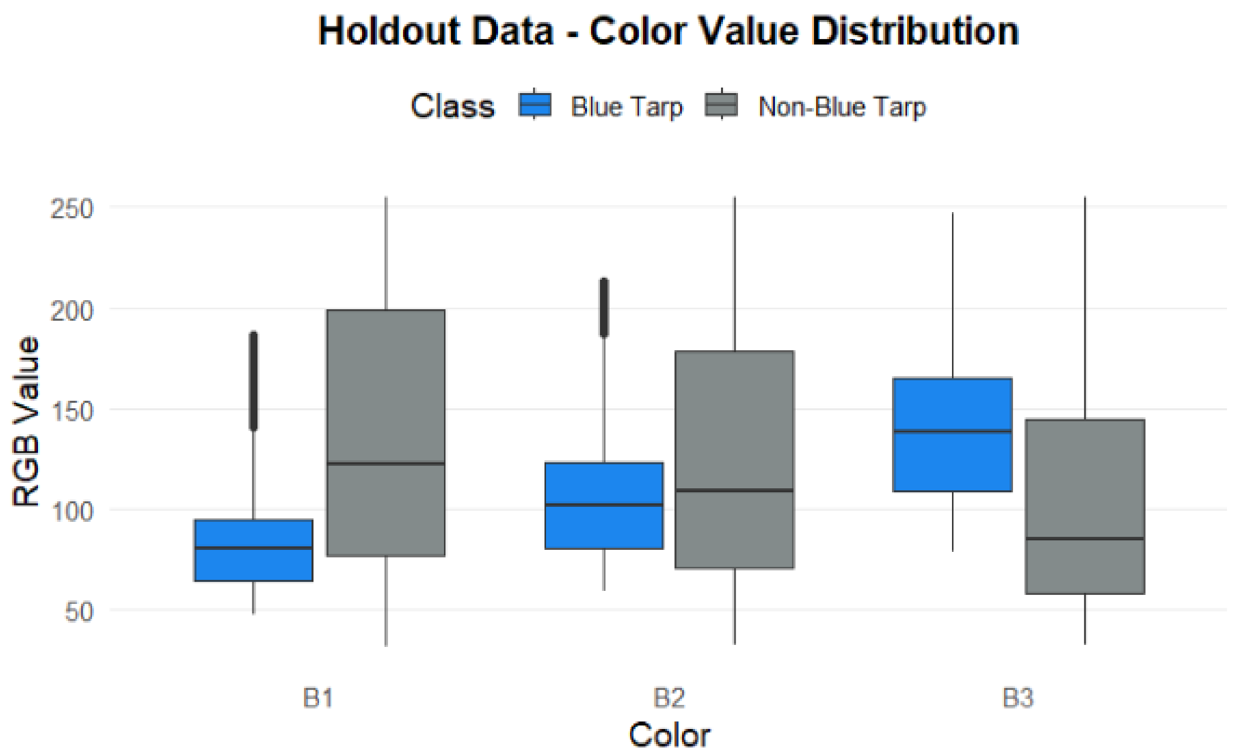
- Used mean comparisons of each channel by class

Improved Approach

- Switched to boxplot distributions to reduce outlier bias

Conclusion

- B1 = Red
- B2 = Green
- B3 = Blue



Preprocessing

Raw Data

Color Standardization

- Mapping B1, B2, and B3 to Red, Green, and Blue

Binary Class Conversion

- Collapsing multiclass labels to Blue Tarp vs. Non-Blue Tarp

Stratified Sampling

- Utilized cross validation with stratified sampling based on class

Normalization

- Scaled the numeric predictors (RGB) for models that margin/distance based (KNN, SVM, etc.)

Modeling Strategy

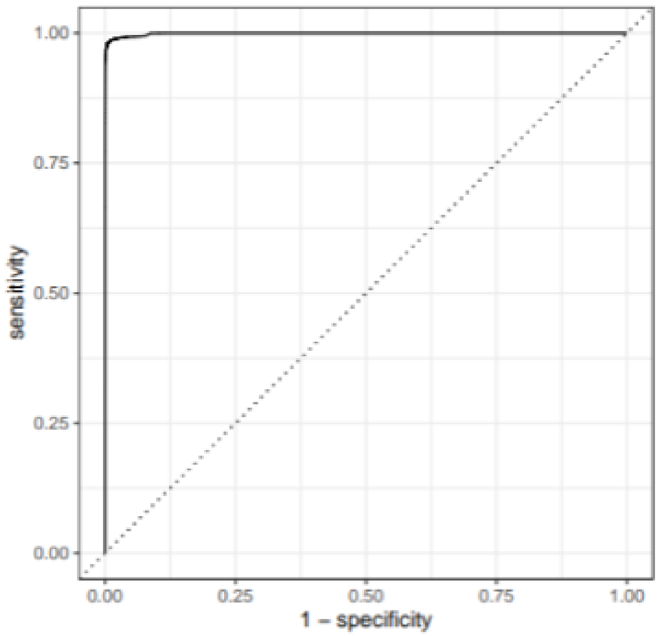
Model	Key Strengths
Logistic Regression	Computationally fast and easy to interpret results
Linear Discriminant Analysis (LDA)	Great at class separation with linear decision boundaries
Quadratic Discriminant Analysis (QDA)	Can handle non-linear decision boundaries, handles class variance differences
K-Nearest Neighbors (KNN)	Non-parametric approach
Support Vector Machine (Linear)	Excellent generalization, effective in high dimensions

- Each model will undergo cross-validation with stratified sampling to account for the stark class imbalance present in both training and holdout data sets.
- For models with tunable parameters, F1-score will be used as the parameter of interest to maximize.
- Should a model have issues with training or holdout performance, threshold tuning will take place to potentially aid the model further.
- Model performance will also be compared from training to holdout data sets to see if a particular model might fall out of favor from signs of overfitting.

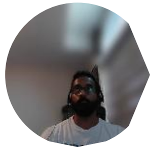
Support Vector Machines (SVM)

- **Goal**:
  - Find the optimal hyperplane that separates Blue Tarp vs. Non-Blue Tarp pixels
- **Process**:
  - We applied 5-fold cross-validation on the training set (stratification by class)
  - Fit the model to each fold using fit_resamples()
  - Collected standard classification metrics
- **Takeaway**:
  - The linear SVM was exceptional in recall and precision
  - Highly reliable in locating blue tarps, with minimal false positives

Metric	ROC AUC	Accuracy	Recall	Precision	F1
Estimate	0.9991	0.9919	0.9919	0.9999	0.9959



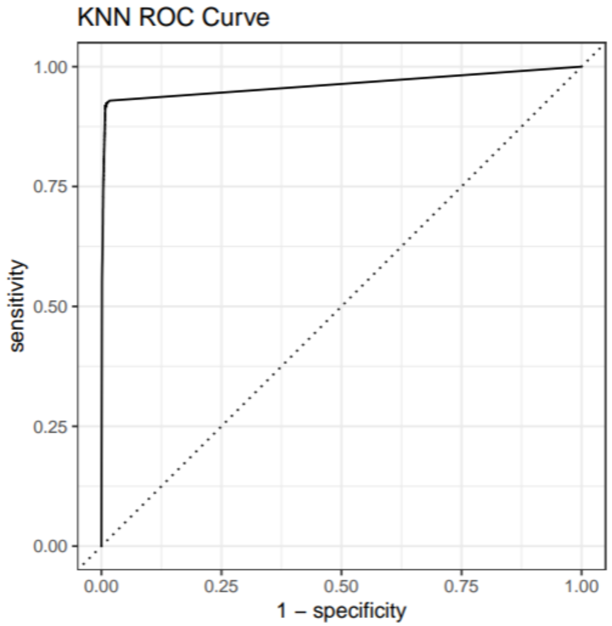
	Predicted Non-Blue	Predicted Blue
Actual Non-Blue	1,973,606	16,083
Actual Blue	223	14,251



K-Nearest Neighbors (KNN)

- **Goal**:
  - Find the optimal number of neighbors that separates Blue Tarp vs. Non-Blue Tarp pixels
- **Process**:
  - We applied 10-fold cross-validation on the training set (stratified by class)
  - Fit the model to each fold using fit_resamples()
  - Tuned the model's number of neighbors to select the best number of neighbors by F1 score
    - Best # of neighbors = 11
  - Picked a threshold to maximize F1 score
    - Threshold: 0.49
  - Collected standard classification metrics against the training performance and holdout performance
- **Takeaway**:
  - The KNN model achieved good training performance but had to tune the threshold due to poor holdout test performance
  - Good accuracy, but poor precision when applied to the holdout set

Metric	ROC AUC	Accuracy	Recall	Precision	F1
Estimate	0.9618	0.9927	0.8379	0.4969	0.6239

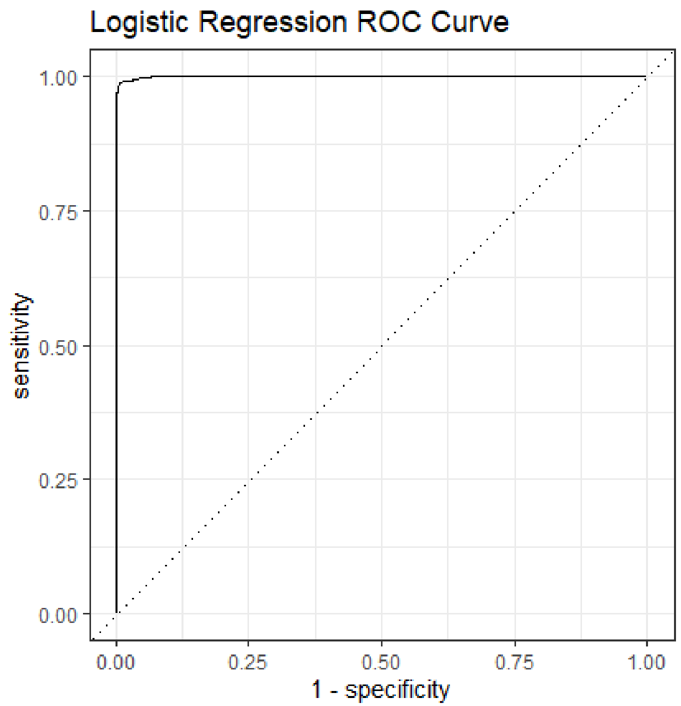


	Predicted Non-Blue	Predicted Blue
Actual Non-Blue	1,977,411	12,278
Actual Blue	2,345	12,129

Logistic Regression

- **Goal:**
  - Develop a sigmoid function that allows for the binary classification of pixels as Blue Tarps or Non-Blue Tarps
- **Process:**
  - We applied 10-fold cross validation with the training set using stratified sampling to account for the class imbalance
  - The model was fit to each fold of data to evaluate its performance
  - Holdout data was used to augment the fitted model and generate predictions and classification performance metrics
- **Takeaway:**
  - The model shows good binary classification performance but struggles with precision
  - The class imbalance impairs the model's performance
    - This could be remedied further by selecting another threshold

Metric	ROC AUC	Accuracy	Recall	Precision	F1
Estimate	0.9994	0.9897	0.9882	0.4131	0.5827

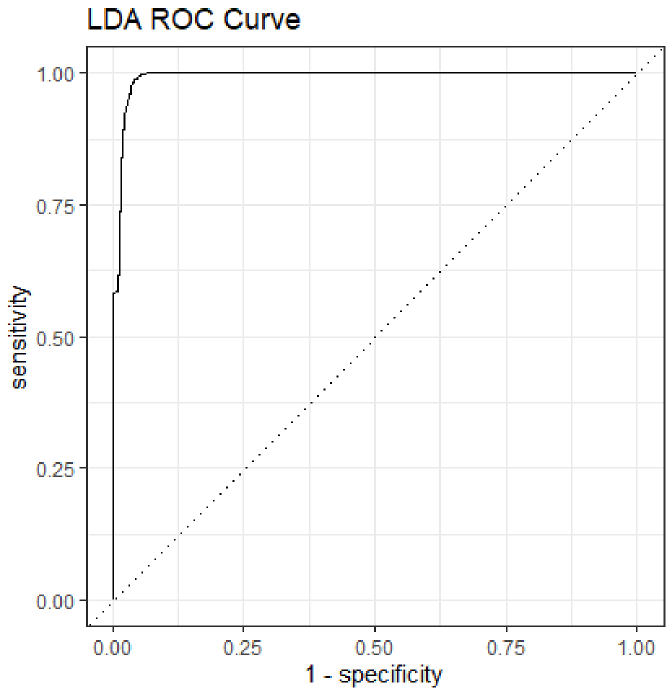


	Predicted Non-Blue	Predicted Blue
Actual Non-Blue	1,969,375	20,314
Actual Blue	170	14,304

Linear Discriminant Analysis (LDA)

- **Goal:**
  - To classify each pixel as either a Blue Tarp or a Non-Blue Tarp through finding a linear combination of features that best separates different classes
- **Process:**
  - We first applied 10-fold cross validation on the training set, then fit the model on the full training data
  - Utilized the holdout data to generate predictions
  - Collected standard classification metrics
- **Takeaway:**
  - The LDA model fared well in ROC_AUC and accuracy but suffered with recall, precision, and F1 score
  - This could be attributed to differing class distributions from training to holdout data sets
    - If the data did not meet the underlying assumptions LDA makes about the data's distribution and covariance of classes, this would make the model suffer

Metric	ROC AUC	Accuracy	Recall	Precision	F1
Estimate	0.9921	0.9817	0.8390	0.2617	0.3990

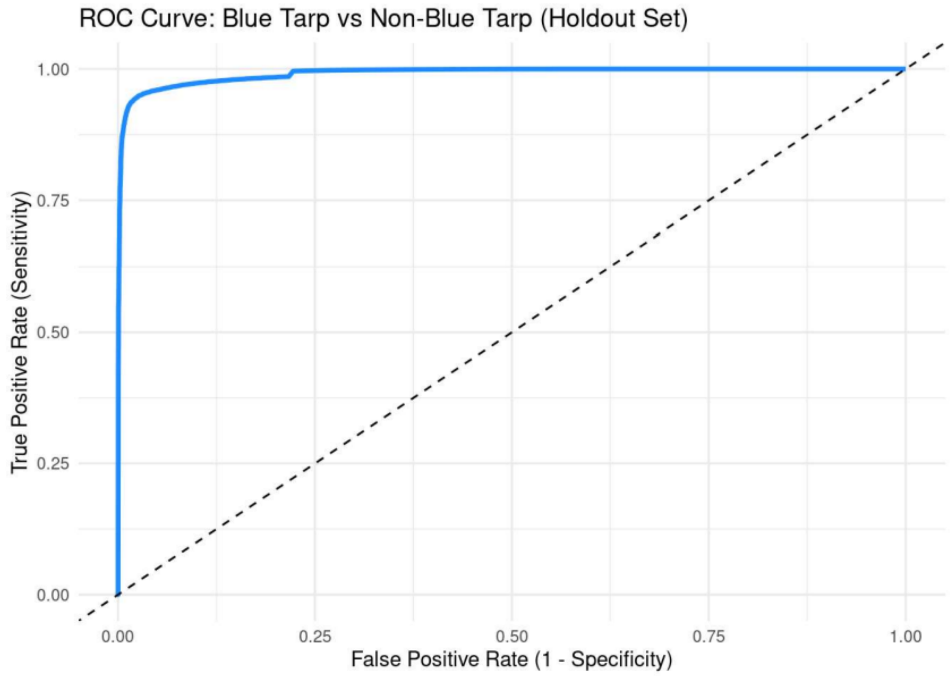


	Predicted Non-Blue	Predicted Blue
Actual Non-Blue	1,955,444	34,245
Actual Blue	2,330	12,144

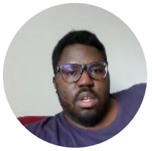
Quadratic Discriminant Analysis (QDA)

- **Goal:**
  - To classify each pixel as either a Blue Tarp or a Non-Blue Tarp
- **Process:**
  - We first applied a 5-fold cross validation on the training set, then fit the model on the full training data
  - Used the holdout data to generate predictions
  - Collected standard classification metrics
- **Takeaway:**
  - The QDA model fared well against the training and holdout sets
  - Seems to successfully distinguish between Blue Tarps and Non-Blue Tarps

Metric	ROC AUC	Accuracy	Recall	Precision	F1
Estimate	0.9915	0.9960	0.6946	0.7736	0.7136



	Predicted Non-Blue	Predicted Blue
Actual Non-Blue	1,986,038	4,419
Actual Blue	3,651	10,055



Holdout Performance Comparison

	Accuracy	Recall	Precision	F1 Score	ROC AUC	Overall
SVM	0.9919	0.9919	0.9999	0.9959	0.9991	Best overall performance – high recall and high F1
KNN	0.9927	0.8379	0.4969	0.6239	0.9618	High accuracy, low F1 – need tuning for class imbalance
Logistic Regression	0.9897	0.9882	0.4131	0.5827	0.9994	Strong, consistent, and interpretable with high AUC
LDA	0.9817	0.839	0.2617	0.399	0.9921	Reliable baseline model
QDA	0.996	0.6946	0.7736	0.7136	0.9915	Balanced metrics with potential risk of overfitting on noisy data