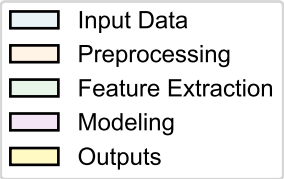


PC-RAI v2.x ML Pipeline Workflow

Random Forest for Rockfall Prediction from Pre-Failure Cliff Morphology



1. INPUT DATA

LiDAR Point Clouds

959 surveys
2017-2025
6 beaches

Event Labels

52,365 events
Volume, location,
elevation, dates

Polygon Shapefiles

1m alongshore bins
× elevation zones
(lower/mid/upper)

2. PREPROCESSING

Step 1:
Survey Selection

Match surveys to
future events
(≥7 day gap)

Step 2a:
Subsample

50cm voxel grid
~10M → ~400K pts

Step 2b:
Compute Normals

CloudComPy MST
Westward oriented

scripts/
01_identify_surveys.py

3. FEATURE EXTRACTION

Step 2c: Extract Point-Level Features

Slope • Roughness (small/large) • Height • Eigenvalues (linearity, curvature) • 9 features per point

02_extract_features.py

4. SPATIAL AGGREGATION

Step 3:
Bin to Polygons

1m alongshore
× 3 elevation zones
= polygon-zones

Aggregate Statistics

mean, std, min, max,
p10, p50, p90
→ 63 features per zone

03_aggregate_polygons.py

5. TRAINING DATA ASSEMBLY

Step 4: Label Polygon-Zones

Cases (1): Future rockfall
Controls (0): No failure
Balanced 1:1 ratio

Final Dataset

72,782 samples
36,391 cases
42 features

04_assemble_training_data.py

6. MODEL TRAINING & EVALUATION

Step 5: Train RF

Random Forest
n_estimators=100
class_weight=balanced

Cross-Validation

Leave-one-year-out
Leave-one-beach-out
5-fold CV

Performance

Temporal CV:
AUC-ROC: 0.701
AUC-PR: 0.667

05_train_model.py

Trained Models

rf_model.joblib

Feature Importance

Ablation Results

Diagnostic Plots

ROC/PR Curves