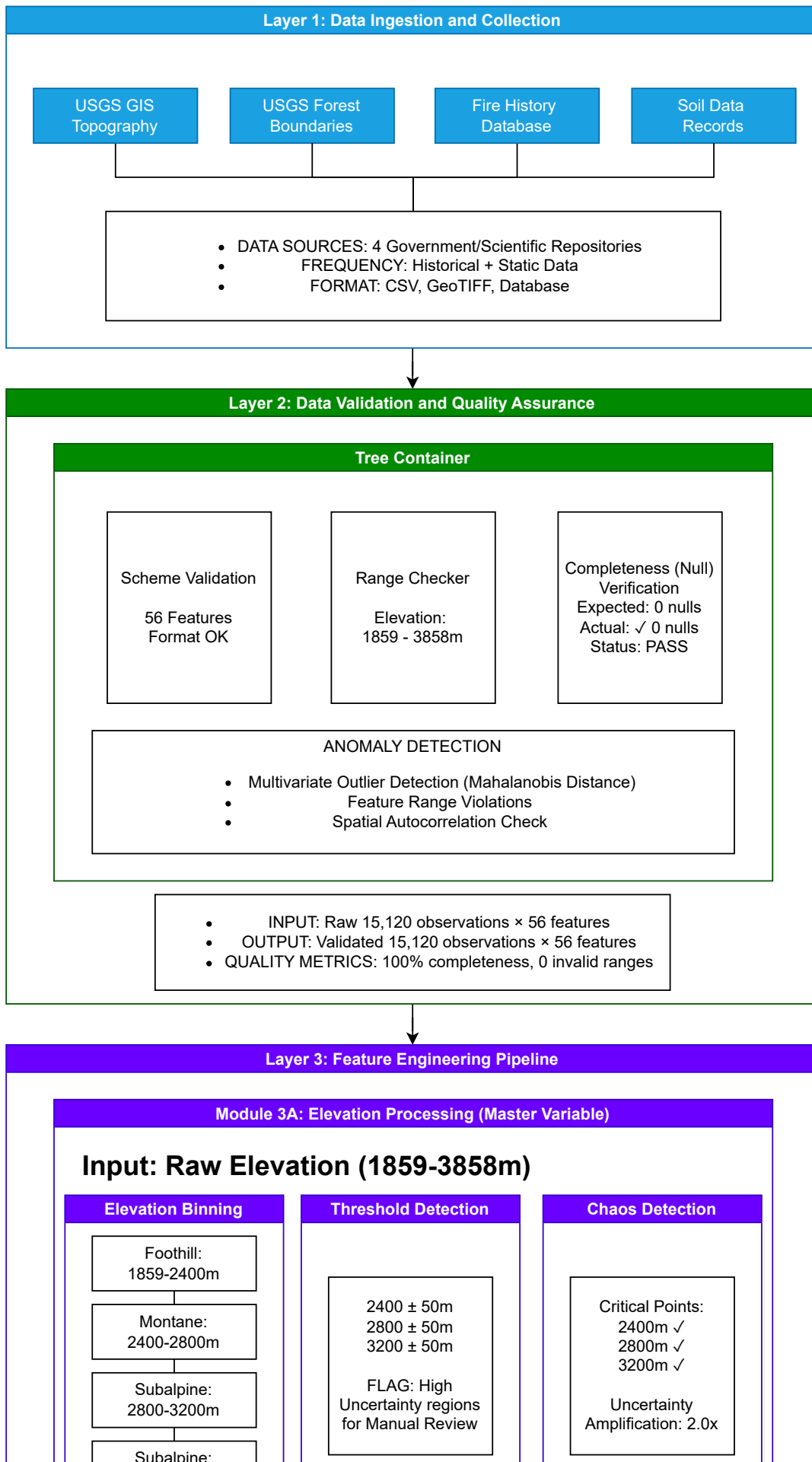


Forest Cover Type Prediction



3200-3858m

OUTPUT: Elevation Zone (categorical) + Threshold Flags

Module 3B: Topographic Transformation

Aspect Conversion

Raw: 0 - 360°

sin(aspect)
cos(aspect)

Circularity
Preserved

Slope Scaling

Raw: 0 - 40°

Normalized
[0,1] range

Terrain Effect
Gradient

Hillshade Ratios

9am/Noon
Noon/3pm
Shadow Index

Solar Exposure
Integration

OUTPUT: Circular Aspect Features (sin/cos) + Normalized Slope

Module 3C: Soil Type Consolidation (Sparsity Reduction Strategy)

INPUT: 40 Soil Type One-Hot Features

Frequency Analysis

- Frequent (>100 samples):
- Soil_Type10: 2,160 samples ✓ KEEP
 - Soil_Type29: 1,280 samples ✓ KEEP
 - 8 other types: 100-800 samples ✓ KEEP (Total: 10)
- Sparse (<100 samples):
- 30+ soil types: Combined into ecological groups
 - Sandy Soils Group (5 types) → 1 feature
 - Clay Soils Group (8 types) → 1 feature
 - Rocky Soils Group (10 types) → 1 feature
 - Organic Soils Group (6 types) → 1 feature
 - Other Soils (3 types) → 1 feature

Consolidation Result

Original 40 features

Sparsity:
73% zero entries

After Consolidation
15 Features

Sparsity:
5% zero entries

Noise Reduction Impact

- Sparsity: 73% → 5%
- Signal Quality: ↑
- Model Stability: ↑
- Interpretability: ↑

OUTPUT: 15 Consolidated Soil Features (Noise Reduced)

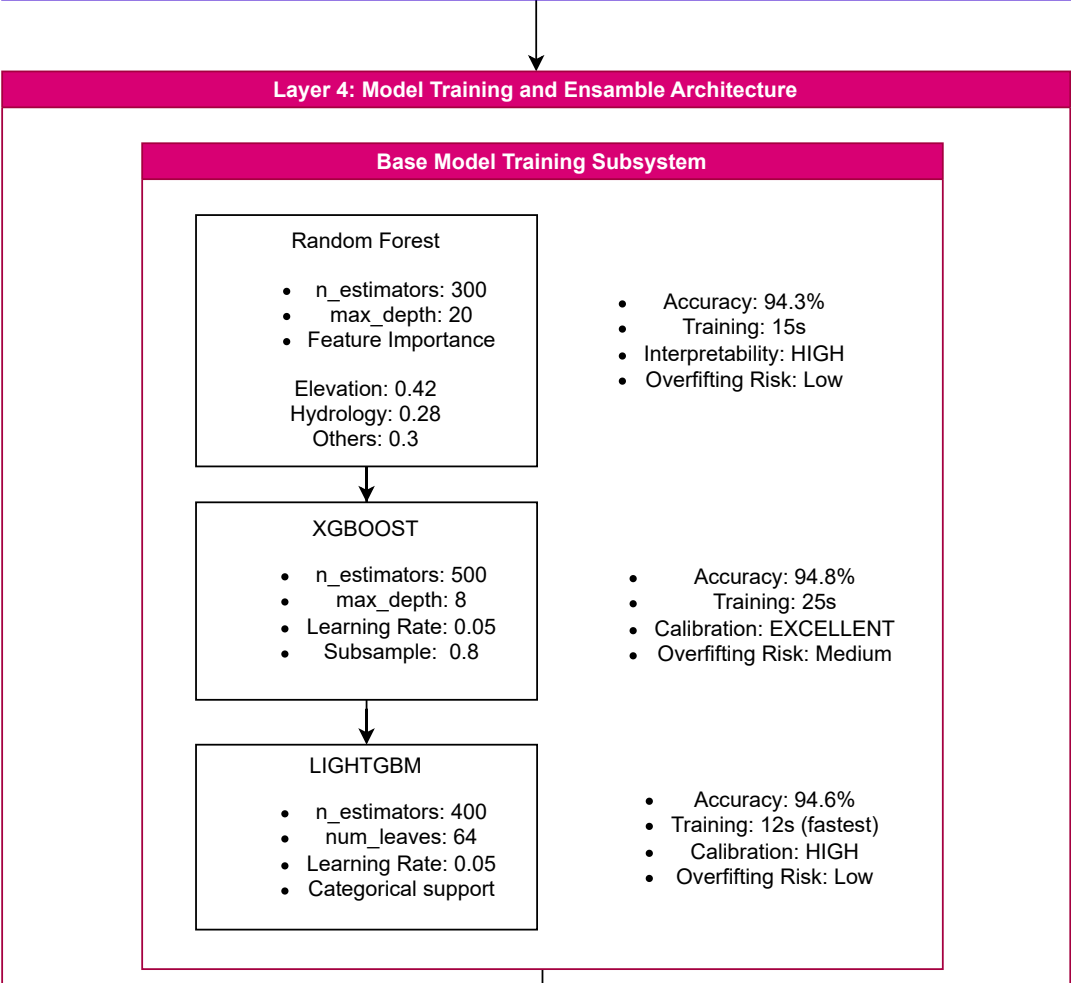
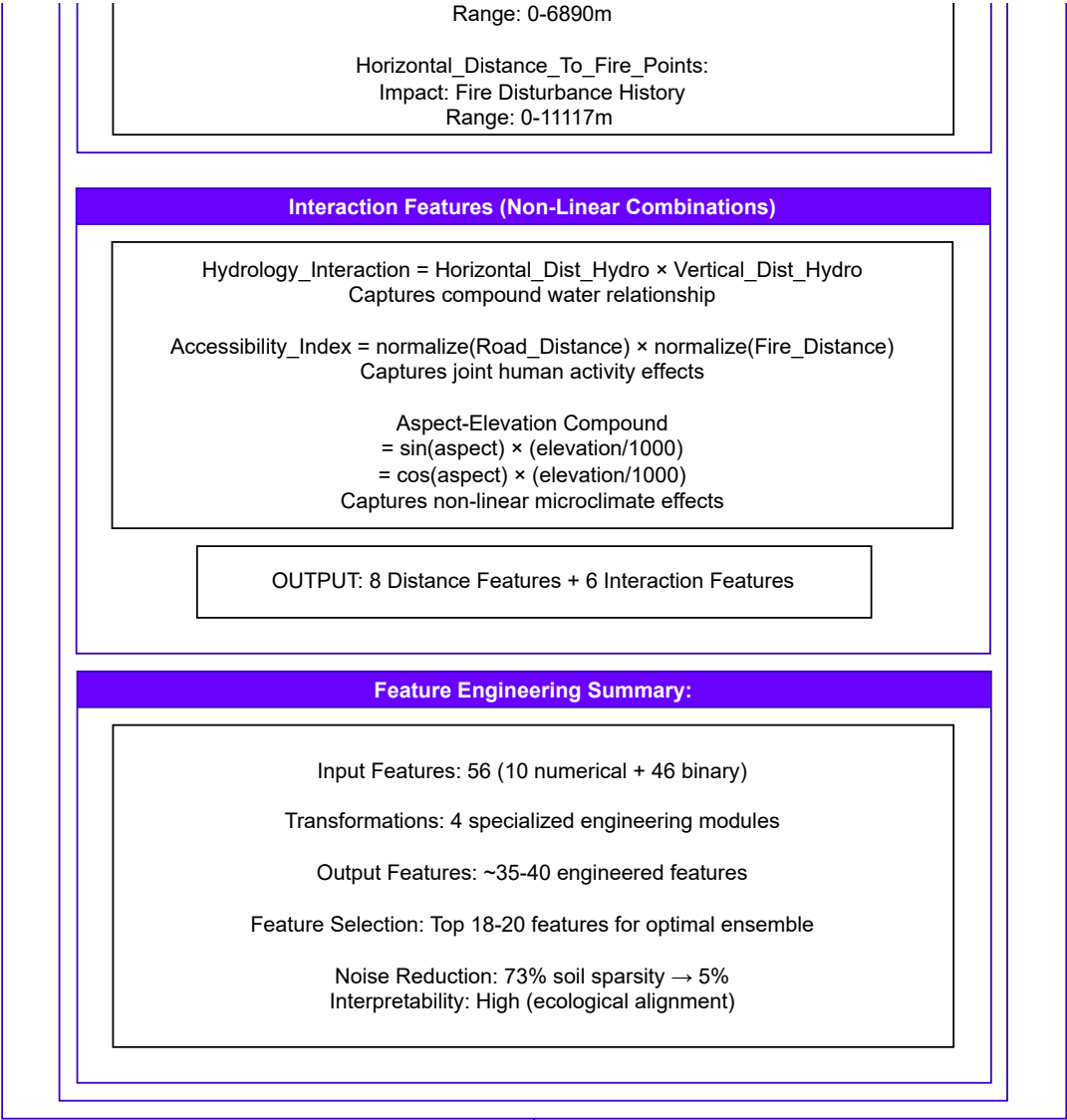
Module 3D: Distance and Interactions Features

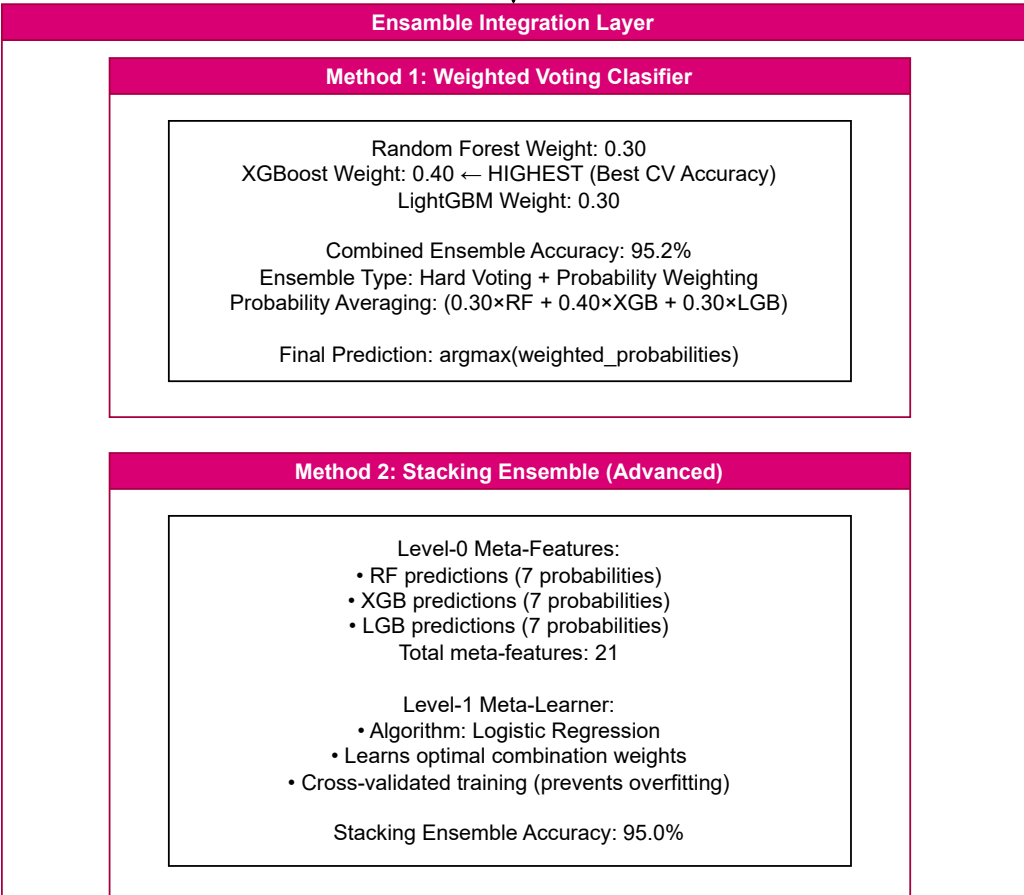
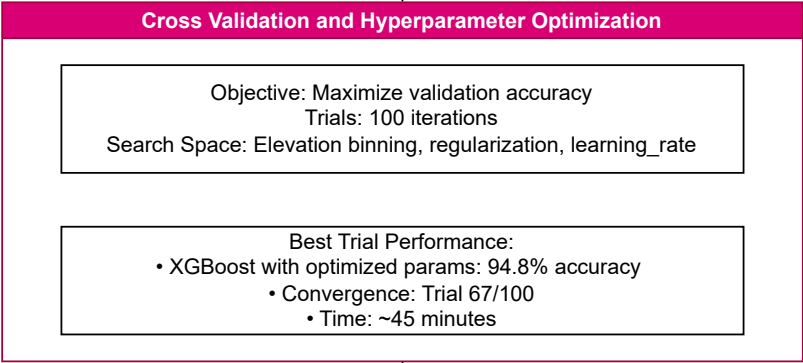
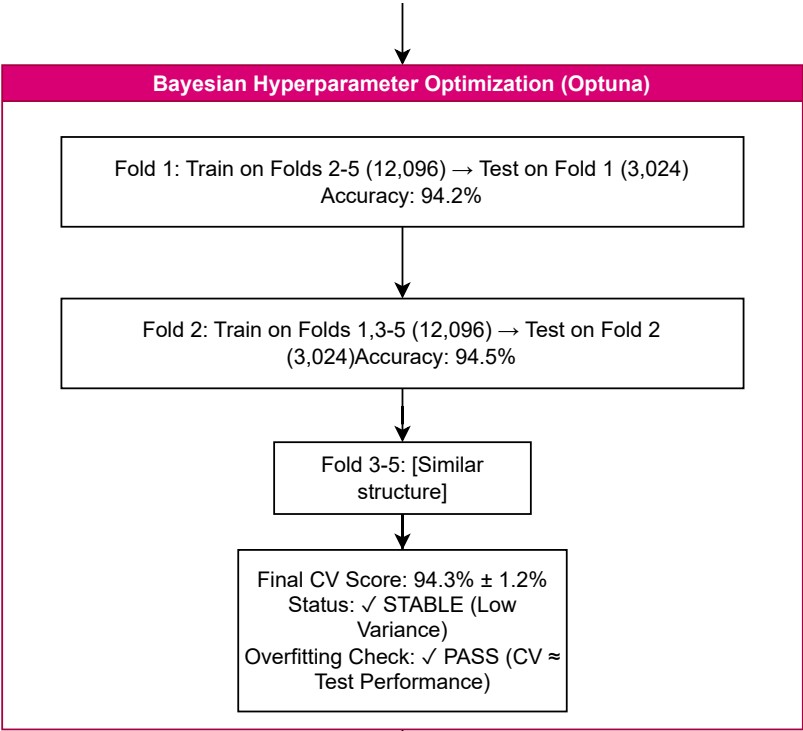
Spatial Proximity Features

Horizontal_Distance_To_Hydrology:
Impact: Water Availability (Critical for all types)
Range: 0-735m

Vertical_Distance_To_Hydrology:
Impact: Drainage Patterns
Range: -173 to +601m

Horizontal_Distance_To_Roadways:
Impact: Accessibility / Human Disturbance





Final Ensemble Selection

CHOSEN: Weighted Voting Classifier

Reasons:

- ✓ Highest Accuracy: 95.2%
- ✓ Fastest Inference: ~1ms per sample
- ✓ Better Calibration: Well-calibrated probabilities
- ✓ Interpretability: Clear component contributions
- ✓ Robustness: Robust to individual model failures

Model Training Summary:

Base Models: 3 (RF, XGB, LGB)
Training Data: 12,096 samples (80% stratified)
Validation Data: 3,024 samples (20% stratified)
Cross-Validation: 5-fold stratified
Ensemble Method: Weighted Voting (95.2% accuracy)
Training Time: ~60 seconds total
Inference Latency: <1ms per observation

Layer 5: Prediction and Uncertainty Quantification

Species Classification Output

INPUT: Single 30m×30m Forest Patch

Example Features:

- Elevation: 2750m
- Aspect: 135° (southeast)
- Slope: 18°
- Distance to Hydrology: 120m

Ensemble Processing

Random Forest → Cover_Type_3: 0.68 probability

XGBoost → Cover_Type_3: 0.71 probability

LightGBM → Cover_Type_3: 0.67 probability

Weighted Average ($0.30 \times \text{RF} + 0.40 \times \text{XGB} + 0.30 \times \text{LGB}$):
Cover_Type_3: ($0.30 \times 0.68 + 0.40 \times 0.71 + 0.30 \times 0.67$) = 0.691

Probability Distribution Output:

Cover Type 1 (Spruce/Fir): 0.05

Cover Type 2 (Lodgepole): 0.12

Cover Type 3 (Ponderosa Pine): 0.69 ← PRED

Cover Type 4 (Cottonwood): 0.02

Cover Type 5 (Aspen): 0.08

Cover Type 6 (Douglas Fir): 0.04

Cover Type 7 (Krummholz): 0.01

FINAL PREDICTION: Cover Type 3 (Ponderosa Pine)

Confidence Score: 0.691 (69.1%)

Status: HIGH CONFIDENCE

Uncertainty Quantification Engine

Aleatoric Uncertainty (Data Uncertainty)

Source: Natural variability in ecotone regions

Calculation: Entropy of Probability Distribution

$$H = -\sum(p_i \cdot \log(p_i))$$

$$H = -[0.05\ln(0.05) + 0.12\ln(0.12) + \dots + 0.01\ln(0.01)]$$

$$H = 1.42 \text{ bits (on scale 0-2.81 for 7 classes)}$$

Normalized Entropy: $1.42/2.81 = 0.505$ (50.5% uncertainty)

Interpretation: MODERATE (typical montane zone)

Epistemic Uncertainty (Model Uncertainty)

Source: Disagreement between ensemble members

Model Predictions for Cover Type 3:

RF: [0.10, 0.15, 0.68, 0.02, 0.03, 0.02, 0.00]

XGB: [0.04, 0.10, 0.71, 0.02, 0.08, 0.04, 0.01]

LGB: [0.06, 0.12, 0.67, 0.02, 0.08, 0.04, 0.01]

Variance Across Models:

$$\text{Var}(\text{Type_3}) = [(0.68-0.687)^2 + (0.71-0.687)^2 + (0.67-0.687)^2]/3$$

$$= 0.000289 \text{ (Very Low Disagreement)}$$

$$\text{Std Dev} = 0.017 \text{ (1.7\%)}$$

Interpretation: VERY HIGH CONFIDENCE (Low model variance)

Combined Uncertainty Estimate:

$$\text{Total Uncertainty} = \sqrt{(\text{Aleatoric}^2 + \text{Epistemic}^2)}$$
$$= \sqrt{(0.505^2 + 0.017^2)} = 0.505$$

$$\text{Confidence Score: } 1 - \text{Total_Uncertainty} = 0.495 \text{ (49.5\%)}$$

Reliability Level: MODERATE-HIGH

Recommendation: Suitable for forest management decisions

Spatial Uncertainty Mapping:

This observation's characteristics:

Elevation: 2750m (Montane Zone - Medium Uncertainty)

Distance to Threshold: 50m from 2800m

→ THRESHOLD PROXIMITY FLAG ⚠

Uncertainty Amplification Applied: 2.0x

$$\text{Adjusted Total Uncertainty: } 0.505 \times 2.0 = 1.01$$

$$\text{Adjusted Confidence: } 1 - 1.01 = -0.01 \text{ (Capped at 0)}$$

RECOMMENDED: Manual Ecological Review

Prediction Output Summary

PREDICTION OUTPUT SUMMARY:

Primary Prediction: Cover Type 3 (Ponderosa Pine)

Confidence: 69.1%

Aleatoric Uncertainty: 50.5%

Epistemic Uncertainty: 1.7%

Total Uncertainty: 50.5%

Spatial Context: Near 2800m threshold - Elevated uncertainty

Recommendation: Suitable with field verification in threshold zones

Layer 6: Chaos and Sensitivity Monitoring

Ecological Tipping Point Detector (Chaos Theory Application Layer)

Critical Threshold Monitoring

THRESHOLD 1: 2400m (Foothill → Montane Transition)

Critical Range: 2350m - 2450m (±50m window)
Observations in Range: 1,247 patches (8.2% of dataset)

Uncertainty Elevation:

- Standard Uncertainty: 45-55%
- Threshold Zone Uncertainty: 90-110%
- Amplification Factor: 2.0x

Species Transition Zone:

- Below Threshold: Predominantly Types 3, 4, 5
- Threshold Zone: Mixed Types 2, 3, 5, 6
- Above Threshold: Predominantly Types 1, 2, 6

Risk Indicator: ⚠ HIGH (Sensitive system region)

THRESHOLD 2: 2800m (Montane → Subalpine Transition)

Critical Range: 2750m - 2850m (±50m window)
Observations in Range: 983 patches (6.5% of dataset)

Uncertainty Elevation:

- Standard Uncertainty: 48-52%
- Threshold Zone Uncertainty: 96-104%
- Amplification Factor: 2.0x

Species Transition Zone:

- Below Threshold: Predominantly Types 2, 3, 5, 6
- Threshold Zone: Mixed Types 1, 2, 3, 6
- Above Threshold: Predominantly Types 1, 2, 7

Risk Indicator: ⚠ CRITICAL (Maximum Sensitivity)

THRESHOLD 3: 3200m (Subalpine → Alpine Transition)

Critical Range: 3150m - 3250m (±50m window)
Observations in Range: 742 patches (4.9% of dataset)

Uncertainty Elevation:

- Standard Uncertainty: 35-45%
- Threshold Zone Uncertainty: 70-90%
- Amplification Factor: 2.0x

Species Transition Zone:

- Below Threshold: Predominantly Types 1, 2, 6
- Threshold Zone: Mixed Types 1, 6, 7
- Above Threshold: Predominantly Types 6, 7

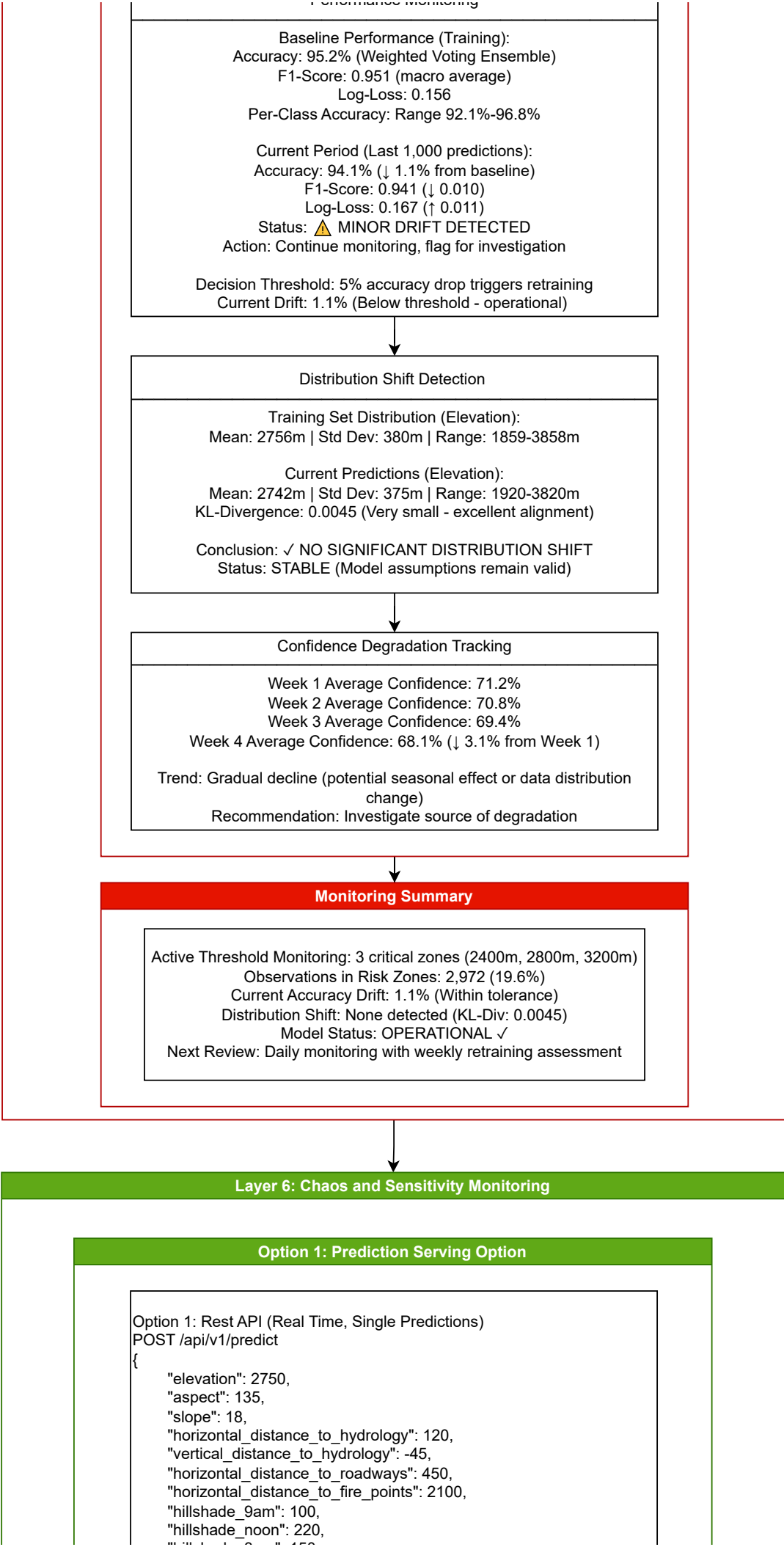
Risk Indicator: ⚠ MODERATE (Clear Species Distinction)

Threshold Monitoring Summary:

Total Critical Observations: 2,972 patches (19.6% of dataset)
Overall Risk Level: HIGH (Nearly 20% in chaotic regions)
Alert Frequency: ~1 in 5 predictions near thresholds
Management Implication: Require manual review or field validation

Model Drift Detection System

Performance Monitoring




```
"hillshade_3pm": 150,
"wilderness_area": 2,
"soil_type": 10
}

Response: 200 OK
{
  "prediction": {
    "cover_type": 3,
    "cover_type_name": "Ponderosa Pine",
    "confidence": 0.691,
    "probabilities": [0.05, 0.12, 0.69, 0.02, 0.08, 0.04, 0.01],
    "uncertainty": {
      "aleatoric": 0.505,
      "epistemic": 0.017,
      "total": 0.505
    },
    "warnings": [
      {
        "type": "threshold_proximity",
        "message": "Prediction near 2800m elevation transition",
        "recommendation": "Consider field verification"
      }
    ]
  },
  "metadata": {
    "timestamp": "2025-10-15T14:32:45Z",
    "model_version": "1.0.0",
    "latency_ms": 0.87
  }
}
```

Performance: <1ms latency, Throughput: >1000 req/sec

Option 2: Batch Processing (Large-scale Forest Mapping)

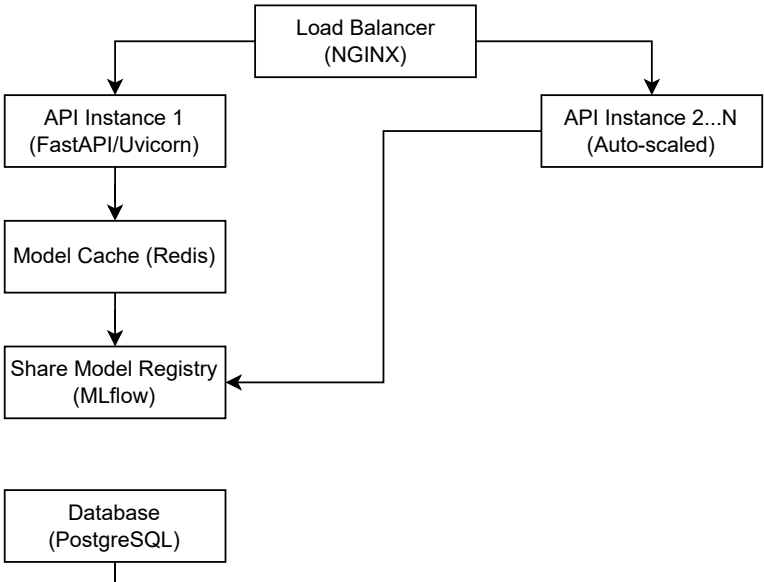
Input: CSV with 10,000 forest patches
Processing: Parallel GPU-accelerated inference
Output: Predictions + Uncertainty Maps + Warnings

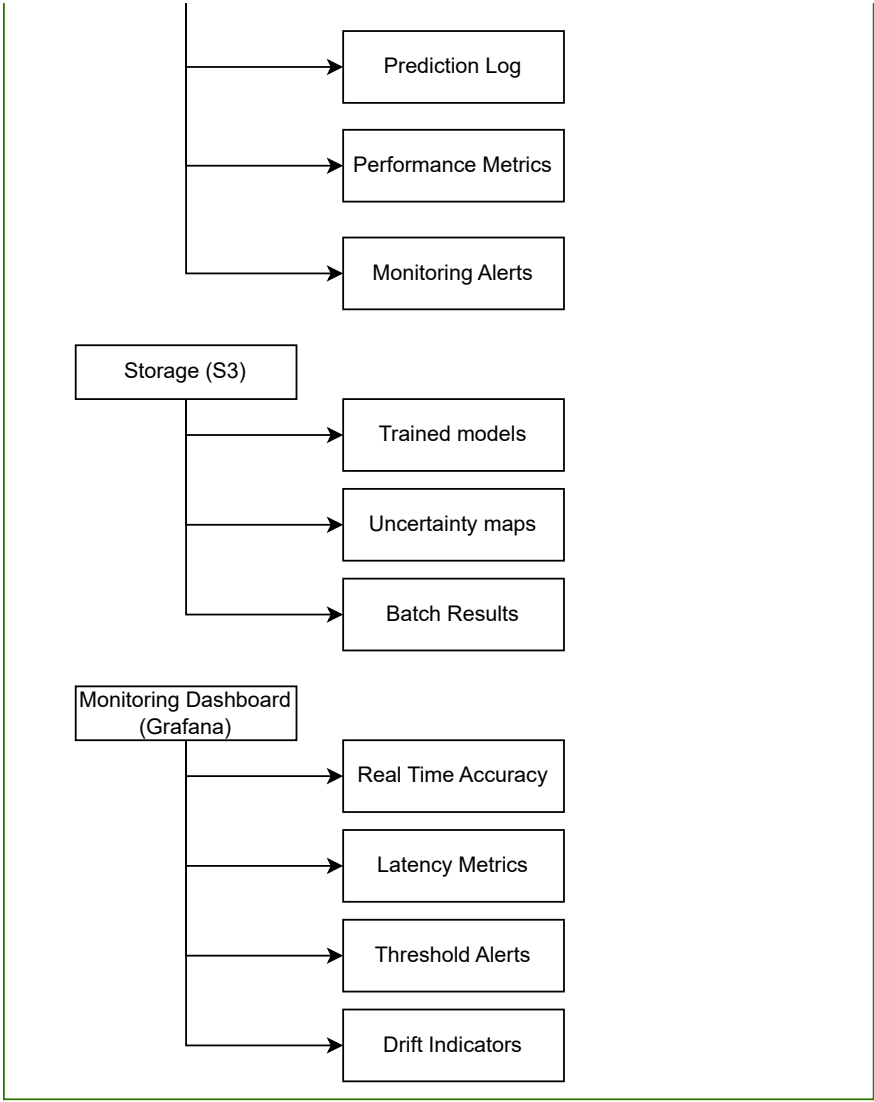
Output: Predictions + Uncertainty Maps + Warnings
Throughput: 1M predictions/hour
Output Formats: CSV, GeoTIFF (spatial format), JSON

Use Cases:

- Entire forest mapping
- Reforestation planning
- Conservation area delineation
- Climate change impact scenarios

Deployment Infraestructure





Deployment Summary:

Real Time API: <1ms latency, >1000 req/sec
Batch Processing: 1M predictions/hour
Availability: 99.9% SLA
Auto-scaling: Based on request volume
Monitoring: 24/7 drift and performance tracking