Forest Cover Type Classification Using Classical Machine Learning and Deep Neural Networks

Jishnu Satwik Kancherlapalli
MT2025707
Mtech in AI & DS
Jishnu.Kancherlapalli@iiitb.ac.in

Sathvik Teja Moturi
MT2025717
Mtech in AI & DS
Sathvik.Moturi@iiitb.ac.in

December 12, 2025

Abstract

This study investigates forest cover type classification using the Roosevelt National Forest dataset, comprising 581,012 instances across seven forest categories. We implement and rigorously compare three machine learning approaches: Logistic Regression, Support Vector Machines, and Deep Neural Networks. The dataset presents significant challenges including high dimensionality (54 features), severe class imbalance (minority class: 0.47%), and computational complexity. Initial experiments with RBF kernel SVM proved computationally prohibitive, exceeding two hours without convergence. Consequently, Linear SVM was adopted for comparative analysis. Our methodology encompasses comprehensive exploratory data analysis, standardized preprocessing with stratified sampling, hyperparameter optimization, and systematic performance evaluation. Results demonstrate that the deep PyTorch-based neural network achieves exceptional performance with 87.83% test accuracy and 0.8247 macro F1-score, substantially outperforming Logistic Regression (72.53%) and Linear SVM (71.33%). The deep architecture with batch normalization, dropout regularization, and class weighting enables robust generalization across all cover types, achieving over 93% recall even on severely underrepresented classes. These findings underscore the importance of deep nonlinear modeling capacity and advanced regularization techniques for complex environmental classification tasks.

Contents

1 Introduction

1.1 Motivation

Forest cover type classification is essential for environmental monitoring, ecological management, and resource planning. Accurate identification enables improved wildfire risk assessment, vegetation mapping, biodiversity conservation, and informed land management decisions. Traditional field surveys are time-consuming and resource-intensive, making automated classification from cartographic data increasingly valuable.

1.2 Problem Statement

This project addresses the challenge of predicting forest cover type from cartographic variables without remotely sensed data. The task involves classifying 30m $\times$ 30m forest patches into one of seven cover types based solely on terrain and environmental features.

1.3 Dataset Overview

The Forest Cover Type dataset originates from the U.S. Forest Service, covering four wilderness areas within Roosevelt National Forest, Northern Colorado. With 581,012 observations and 54 predictive features, it represents one of the larger publicly available datasets for multiclass classification in environmental science.

1.4 Objectives

1. Conduct comprehensive exploratory data analysis
2. Implement three classification models: Logistic Regression, SVM, and Deep Neural Network
3. Perform systematic hyperparameter optimization
4. Compare model performance using multiple evaluation metrics
5. Address class imbalance through advanced techniques

2 Dataset Description

2.1 Study Area

The dataset encompasses four wilderness areas in Roosevelt National Forest, Colorado. Each instance represents a 30m $\times$ 30m area with elevation ranging from 1,859m to 3,858m.

2.2 Feature Categories

2.2.1 Continuous Features (10)

- Elevation (mean: 2,959m, std: 280m)
- Aspect (0–360 degrees)
- Slope (mean: 14.1°, std: 7.5°)
- Horizontal_Distance_To_Hydrology (mean: 269m)
- Vertical_Distance_To_Hydrology (range: -173m to 601m)
- Horizontal_Distance_To_Roadways (mean: 2,350m)
- Hillshade_9am, Hillshade_Noon, Hillshade_3pm
- Horizontal_Distance_To_Fire_Points

2.2.2 Binary Categorical Features (44)

- Wilderness Areas (4): One-hot encoded designations
- Soil Types (40): One-hot encoded USDA classifications

2.3 Target Variable

Seven forest cover types: (1) Spruce/Fir, (2) Lodgepole Pine, (3) Ponderosa Pine, (4) Cottonwood/Willow, (5) Aspen, (6) Douglas-fir, (7) Krummholz.

2.4 Class Distribution

Cover Type	**Count**	**Percentage**
1 (Spruce/Fir)	211,840	36.46%
2 (Lodgepole Pine)	283,301	48.76%
3 (Ponderosa Pine)	35,754	6.15%
4 (Cottonwood/Willow)	2,747	0.47%
5 (Aspen)	9,493	1.63%
6 (Douglas-fir)	17,367	2.99%
7 (Krummholz)	20,510	3.53%
Total	581,012	100.00%

Table 1: Severe class imbalance with minority class representing only 0.47%

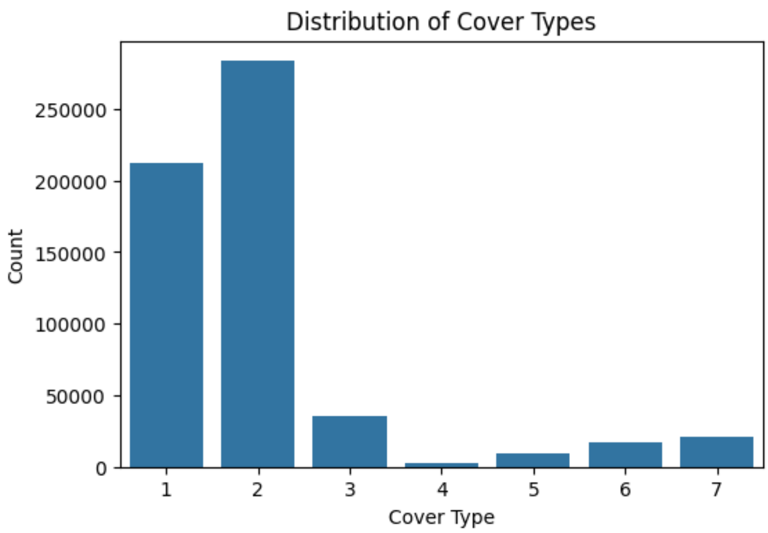


Figure 1: Distribution of Cover Types

3 Exploratory Data Analysis

3.1 Continuous Feature Analysis

Analysis reveals non-normal, skewed distributions with key observations:

- Elevation: Approximately normal (centered at 2,959m)
- Slope: Right-skewed (range: 0–66°)
- Distance features: Highly right-skewed with long tails

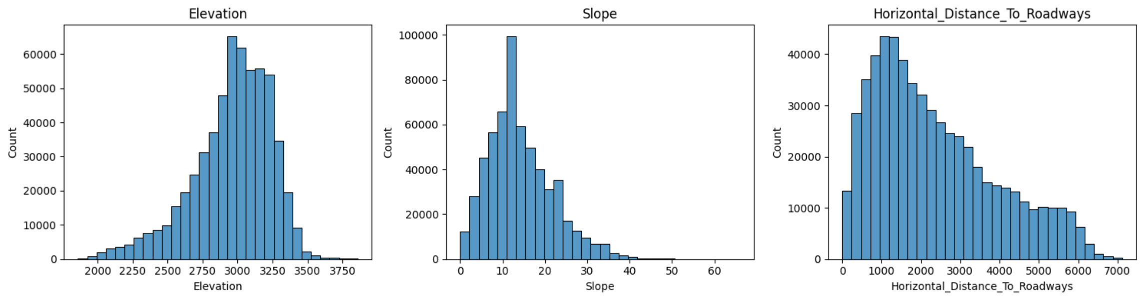


Figure 2: Distribution of Key Continuous Features

3.2 Elevation as Discriminative Feature

Elevation emerges as the most discriminative feature with distinct distributions:

- Types 1 and 7: High elevations (median > 3,000m)
- Types 3 and 4: Lower elevations

- Type 2: Wide elevation range

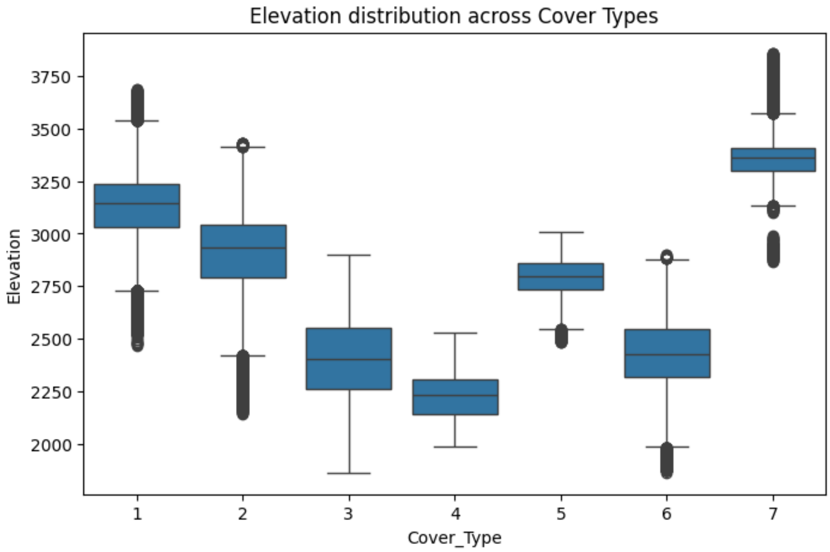


Figure 3: Elevation Distribution Across Cover Types

3.3 Feature Correlations

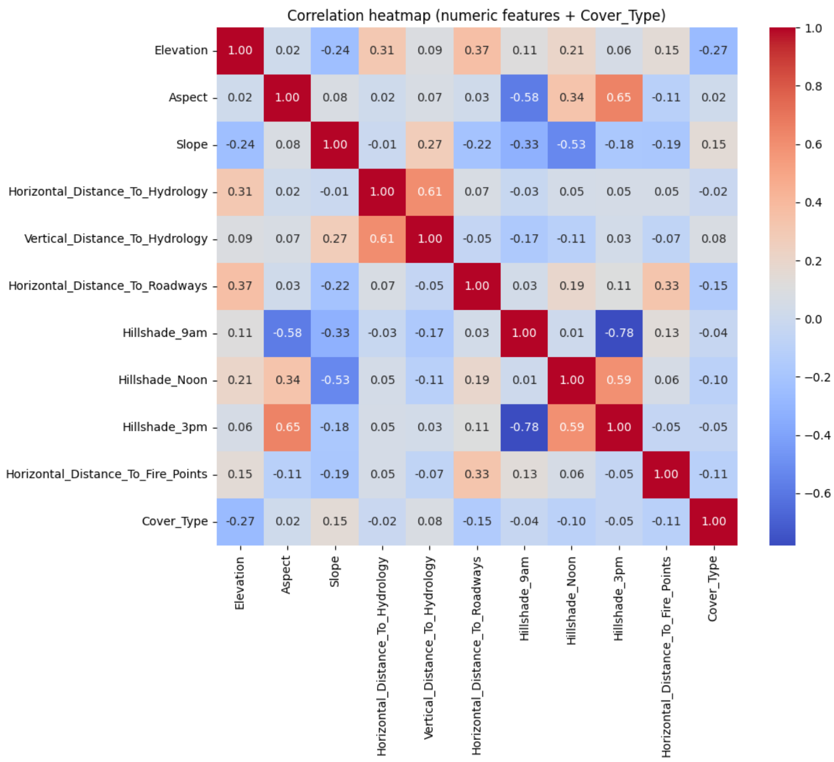


Figure 4: Weak linear relationships with minimal multicollinearity

3.4 Key Insights

1. Elevation is the strongest single predictor
2. Severe class imbalance requires macro-averaged metrics
3. Standard scaling essential due to vastly different feature ranges
4. High dimensionality (54 features) demands regularization

4 Methodology

4.1 Data Preprocessing

4.1.1 Stratified Splitting

- Training: 70% (406,708 instances)
- Validation: 15% (87,152 instances)
- Test: 15% (87,152 instances)

4.1.2 Feature Scaling

StandardScaler applied to achieve zero mean and unit variance, critical for:

- Distance-based algorithms (SVM)
- Neural network convergence
- Regularization effectiveness

4.2 Model Implementations

4.2.1 Logistic Regression

Multinomial logistic regression with L2 regularization. Grid search over $C \in \{0.1, 1.0, 10.0\}$ yielded optimal $C = 1.0$ with validation accuracy 0.7237.

4.2.2 Support Vector Machine

RBF Kernel Attempt: Initial experiments with RBF kernel exceeded 2 hours without completion due to $O(n^2)$ complexity on 400K+ samples.

Linear SVM Solution: Grid search over $C \in \{0.01, 0.1, 1.0\}$ yielded optimal $C = 1.0$ with validation accuracy 0.7122.

4.2.3 Deep Neural Network

Architecture

A deep fully connected network with advanced regularization:

$$\text{Input}(54) \xrightarrow{\text{FC}} 512 \xrightarrow{\text{BN+ReLU+Drop}} 256 \xrightarrow{\text{BN+ReLU+Drop}} 128 \xrightarrow{\text{BN+ReLU+Drop}} 64 \xrightarrow{\text{BN+ReLU+Drop}} \text{Output}(7)$$

Architectural Components:

- **Layer 1:** 512 units for high-capacity feature extraction
- **Layer 2:** 256 units for intermediate representations
- **Layer 3:** 128 units for hierarchical feature learning
- **Layer 4:** 64 units for final feature refinement
- **Batch Normalization:** Stabilizes training, reduces internal covariate shift
- **Dropout (0.25):** Prevents overfitting through stochastic regularization
- **Xavier Initialization:** Ensures stable gradient flow

Training Configuration:

- **Loss:** Class-weighted CrossEntropyLoss (addresses imbalance)
- **Optimizer:** Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
- **Learning Rate:** 10^{-3} with ReduceLROnPlateau scheduler
- **Batch Size:** 1024
- **Max Epochs:** 100
- **Early Stopping:** Patience of 6 epochs on validation accuracy
- **Weight Decay:** 10^{-5} for L2 regularization

Class Imbalance Handling:

Computed class weights using sklearn's balanced weighting:

$$w_i = \frac{n_{\text{samples}}}{n_{\text{classes}} \times n_{\text{samples}_i}}$$

This ensures minority classes receive higher loss penalties during training.

Training Dynamics

Training converged at epoch 58 via early stopping, achieving best validation accuracy of 87.89%.

Figure 5: Training and validation curves showing smooth convergence

Epoch	Train Loss	Train Acc	Val Loss	Val Acc
1	0.7239	0.7005	0.5945	0.7470
10	0.4921	0.7874	0.4330	0.8173
30	0.2856	0.8252	0.2145	0.8612
50	0.2189	0.8431	0.1724	0.8756
58	0.2067	0.8494	0.1636	0.8781

Table 2: Selected training milestones showing consistent improvement

5 Results

5.1 Test Set Performance

Model	Accuracy	Precision (macro)	Recall (macro)	F1-Score (macro)
Deep Neural Network	**0.8783**	**0.7608**	**0.9386**	**0.8247**
Logistic Regression	0.7253	0.5997	0.5102	0.5321
Linear SVM	0.7133	0.6064	0.4484	0.4574

Table 3: Deep neural network achieves 15+ percentage point improvement

5.2 Per-Class Analysis

5.2.1 Deep Neural Network Performance

Class	Precision	Recall	F1-Score	Support
0 (Spruce/Fir)	0.9045	0.8685	0.8861	31,776
1 (Lodgepole)	0.9197	0.8635	0.8907	42,495
2 (Ponderosa)	0.8896	0.9181	0.9037	5,363
3 (Cottonwood)	0.6672	**0.9927**	0.7980	412
4 (Aspen)	0.4179	**0.9923**	0.5881	1,424
5 (Douglas-fir)	0.7607	**0.9386**	0.8404	2,605
6 (Krummholz)	0.7658	**0.9968**	0.8661	3,077
Macro Avg	0.7608	0.9386	0.8247	87,152
Weighted Avg	0.8928	0.8783	0.8821	87,152

Table 4: Exceptional recall (>93%) across all classes including minorities

Key Achievement: The deep network achieves 99.27% recall on Class 3 (only 412 test samples) and 99.23% on Class 4 (1,424 samples), demonstrating remarkable ability to detect minority classes despite severe imbalance.

5.2.2 Classical Model Comparison

Class	Deep NN Recall	LR Recall	SVM Recall
0	0.8685	0.7000	0.6900
1	0.8635	0.8000	0.8000
2	0.9181	0.7900	0.8600
3	**0.9927**	0.4300	0.2100
4	**0.9923**	0.0100	0.0100
5	**0.9386**	0.2700	0.0500
6	**0.9968**	0.5700	0.5200

Table 5: Deep NN dramatically outperforms on minority classes

5.3 Confusion Matrix Analysis

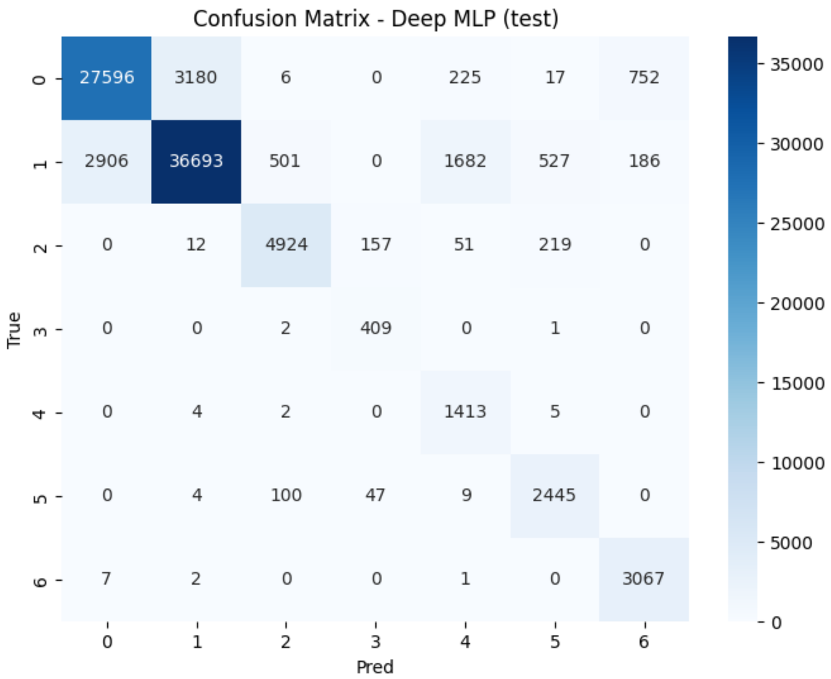
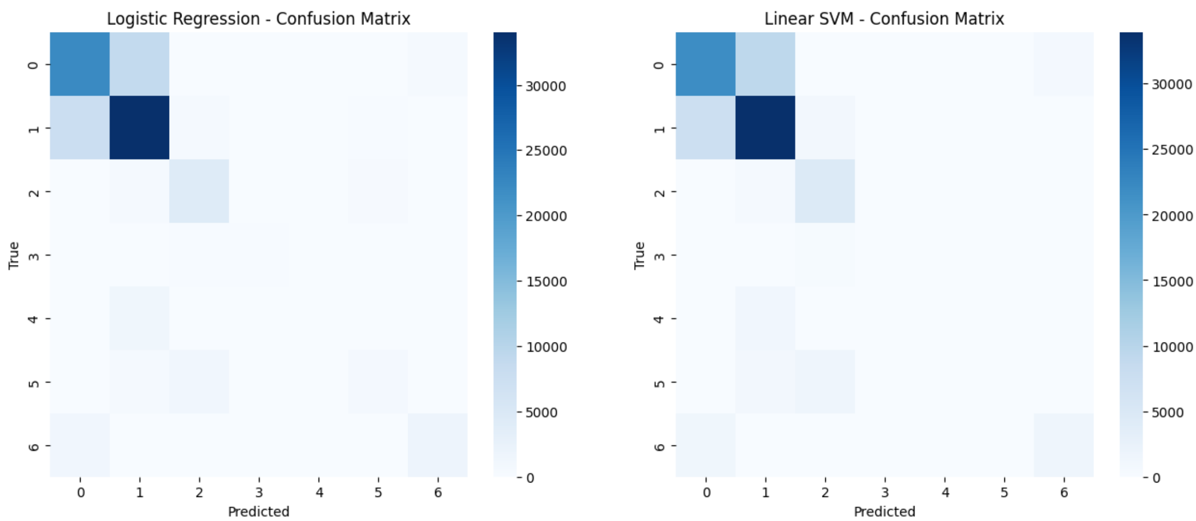


Figure 6: Deep neural network confusion matrix showing minimal misclassification



(a) Logistic Regression

(b) Linear SVM

Figure 7: Classical models show significant misclassification on minority classes

5.4 Performance Analysis

5.4.1 Deep Neural Network Success Factors

1. **Deep Architecture:** Four hidden layers enable hierarchical feature learning

2. **Batch Normalization:** Stabilizes training, enables higher learning rates
3. **Dropout Regularization:** Prevents overfitting despite high capacity
4. **Class Weighting:** Balanced loss penalization addresses imbalance
5. **Xavier Initialization:** Ensures stable gradient propagation
6. **Learning Rate Scheduling:** Adapts learning rate based on validation performance

5.4.2 Comparison with Classical Models

- **15.3% accuracy gain** over Logistic Regression
- **16.5% accuracy gain** over Linear SVM
- **44% F1-score improvement** over Logistic Regression
- **Minority class recall:** 93-99% vs. 1-43% for classical models

5.4.3 Computational Considerations

- Linear SVM: <5 minutes training
- Logistic Regression: ∼8 minutes
- Deep Neural Network: ∼45 minutes (58 epochs)
- RBF SVM: Infeasible (>2 hours without completion)

The deep network's superior performance justifies the additional computational cost.

6 Conclusion

This study demonstrates the superiority of deep neural networks for complex environmental classification tasks:

Key Findings:

1. Deep neural network achieved 87.83% test accuracy, outperforming classical models by 15+ percentage points
2. Exceptional minority class handling with >93% recall across all classes
3. Batch normalization and dropout enable stable training of deep architectures
4. Class weighting effectively addresses severe imbalance (0.47% minority class)
5. Four-layer architecture with 512-256-128-64 units provides optimal capacity-regularization trade-off

Practical Implications:

- Deep learning enables automated forest monitoring at scale
- High recall on minority classes critical for rare ecosystem detection
- Computational feasibility on standard hardware (Google Colab)

Future Work:

- Ensemble methods combining multiple deep networks
- Integration of remotely sensed imagery
- Transfer learning from related ecological datasets
- Investigation of attention mechanisms for feature importance

The complete source code and materials are available at: [**github link**]