Slope Stability Prediction Using Machine Learning Approaches

Alaguvalliappan Thiagarajan[1]

[1]Affiliation not available

December 21, 2023

Slope Stability Prediction Using Machine Learning Approaches Considering Climate Change

Alaguvalliappan Thiagarajan[1], Dr. Suguang Xiao[2]

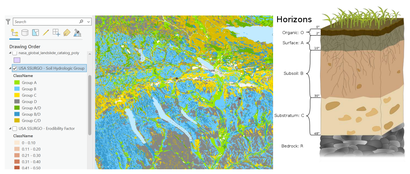
[1]Department of Mathematics, Department of Statistics, University of Florida, [2]Department of Civil and Environmental Engineering, Clarkson University

Motivation

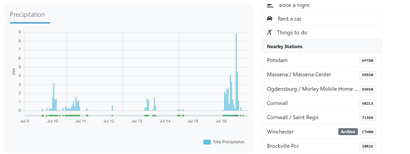
Infrastructure, communities, and ecosystems will all be more vulnerable to landslides in the future due to the increased frequency of extreme weather events brought on by climate change. We need to have accurate, wide scale systems in place to identify landslide susceptible areas, especially since landslides already cost the US 2-4 billion in damages per year

Data

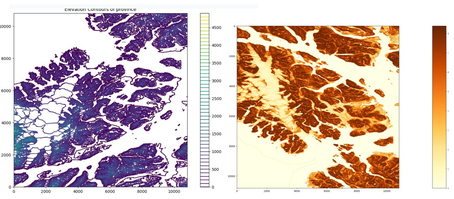
SSURGO Soil Parameters



Meteostat Precipitation Data



USGS 10m Elevation Dataset



Labels from NASA Landslide Inventory and Random Generation

Methodology-Continued

Cross Validation (CV)

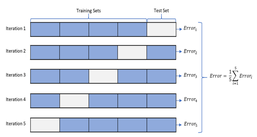
- Models exhibited high variance with different test sets.
- Cross-validation was employed to compare the general performance between models
- Cross validation ensured models were not overfit and could generalize well

Random Seed 42 Results

		precision	recall	f1-score	support
	0	0.90	0.83	0.86	53
	1	0.85	0.91	0.88	55
	accuracy			0.87	108
	macro avg	0.87	0.87	0.87	108
	weighted avg	0.87	0.87	0.87	108

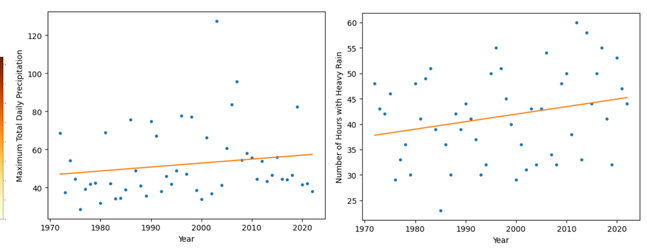
Random Seed 43 Results

		precision	recall	f1-score	support
	0	0.94	0.97	0.95	60
	1	0.96	0.92	0.94	48
	accuracy			0.94	108
	macro avg	0.95	0.94	0.94	108
	weighted avg	0.94	0.94	0.94	108



Future Weather Prediction

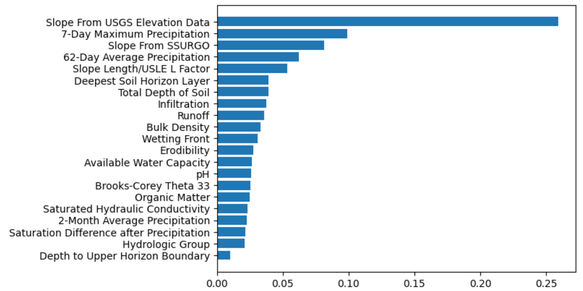
- To predict the maximum daily rainfall per year and the number of hours with heavy rain per year, we collect the data for these parameters in each location from 1972 to 2022
- We created a line of best fit from which we apply m*(2073)+b to get our prediction for the parameters in 2073



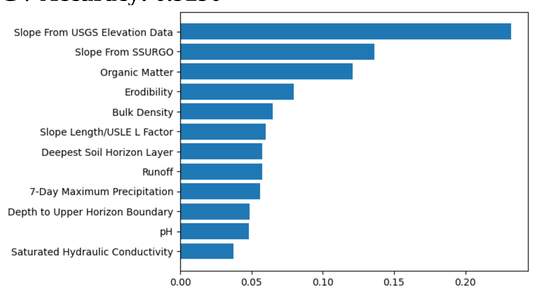
Feature Selection

Recursive Feature Selection (RFE)

- Recursive Feature Elimination trains the model after iteratively removing features to find the subset of features with the best performance.



Random Forest Full Model
CV Accuracy: 0.9290

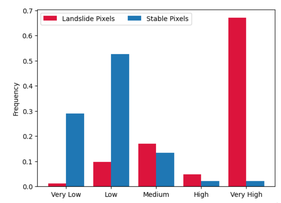


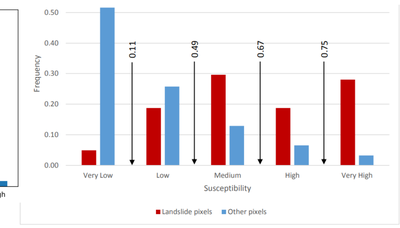
Random Forest Features Chosen by RFE
CV Accuracy: 0.9349

Results

Comparison of our Model to NASA LHASA 1.1

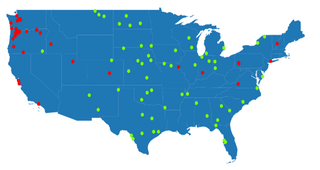
- Due to underreporting, stability of pixels outside the landslide inventory is uncertain.
- As a result, LHASA 1.1 chooses to predict the susceptibility of both landslide and non-landslide pixels instead of using standard classification metrics to evaluate the model.
- To convert our Random Forest's results into a probabilistic score, we count the percentage of decision trees that classify an area as unstable
- Our model generally outperforms LHASA 1.1, classifying around **70** percent of landslide points as having high or very high susceptibility compared to LHASA's nearly **50** percent.



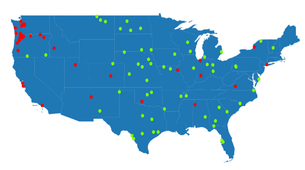


Test Set Performance

While the performance differs depending on the choice of test set, the mean test accuracy is approximately **92** percent. From the images below, we can see that our model performs quite well but is biased toward false positives



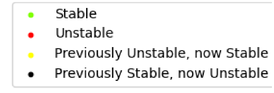
Actual Values in Test Set

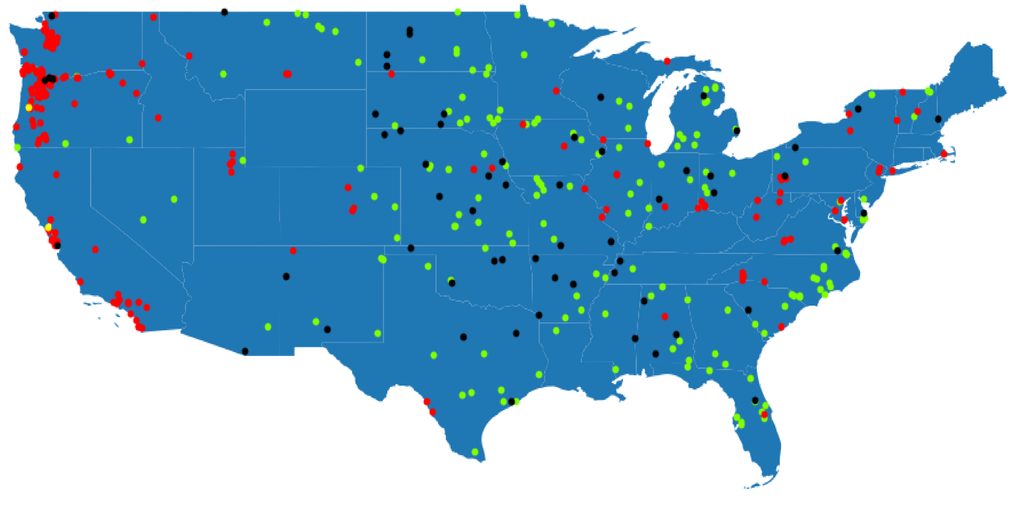


Predicted Values in Test Set

2073 Landslide Susceptibility Forecasting

Random Forest prediction of Landslide Susceptibility in 2073 using predicted values for Maximum Daily Rainfall for each location in 2073 and the number of hours with heavy rain for each location in 2073

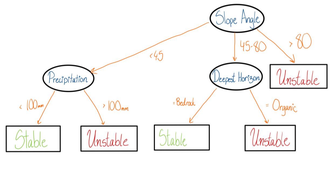


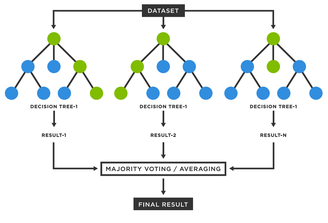


Methodology-Model Choice

Random Forest

- A random forest is a machine learning model that aggregates the results of decision trees
- Each decision tree is trained on a random subset of features





Conclusions & Future Work

- Limited number of samples prevented deep neural networks from converging, and the complexity of the problem made linear regression inadequate.
- Support Vector Machines and ensemble models performed well, with Random Forest performing the best.
- Random Forest achieved performance surpassing LHASA 1.1 in the United States for landslide susceptibility prediction.
- Our data collection methods require validation from empirical data collected from landslide case sites.
- Our dataset focuses on rainfall induced landslides rather than other triggers so more data is needed to account for parameters like seismic activity.

References

ACKNOWLEDGMENT
This work is supported by National Science Foundation under Grant Nos, OAC-2244049.