Slope Stability Prediction Using Machine Learning Approaches Considering Climate Change

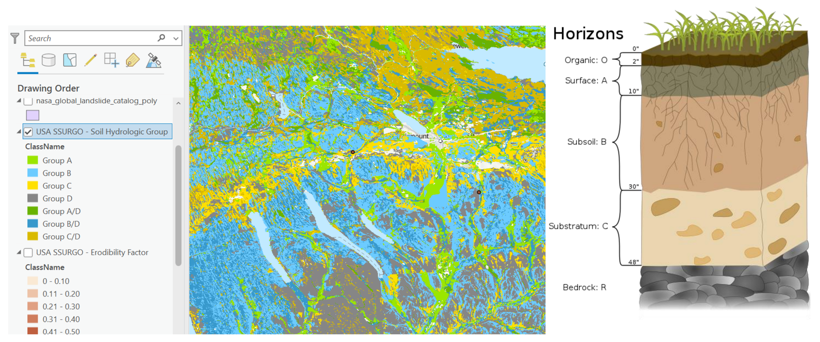
Alaguvalliappan Thiagarajan[1], Dr.Suguang Xiao[2]
[1]Department of Mathematics, Department of Statistics, University of Florida, [2]Department of Civil and Environmental Engineering, Clarkson University

Motivation

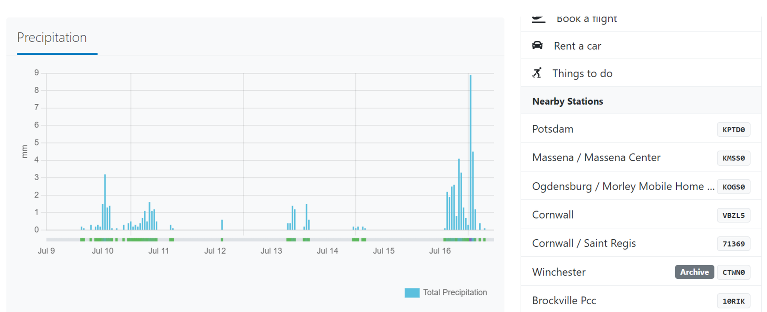
Due to the rise in extreme weather conditions due to climate change, landslides will pose an escalating threat in the future, affecting infrastructure, communities, and ecosystems. In order to combat this issue, we need to have accurate, wide scale systems in place in order to identify landslide susceptible areas.

Data

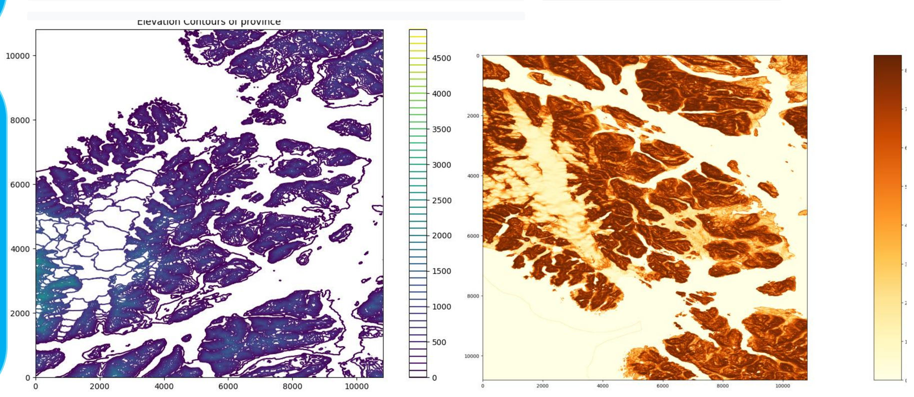
SSURGO Soil Parameters



Meteostat Precipitation Data



USGS 10m Elevation Dataset

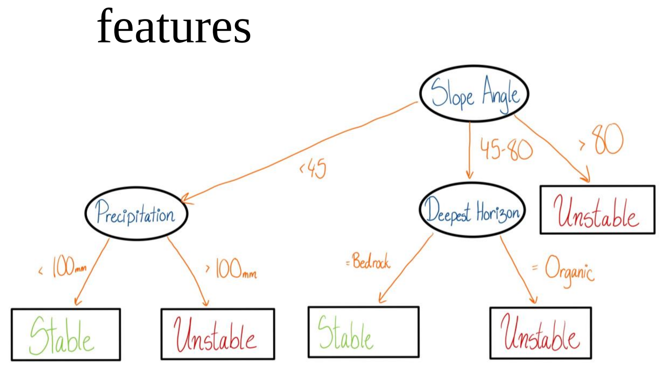


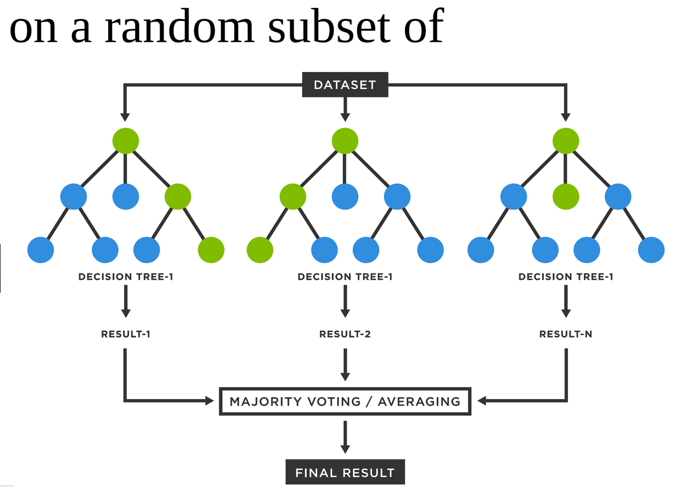
Labels from NASA Landslide Inventory and Random Generation

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





Methodology-Continued

Cross Validation

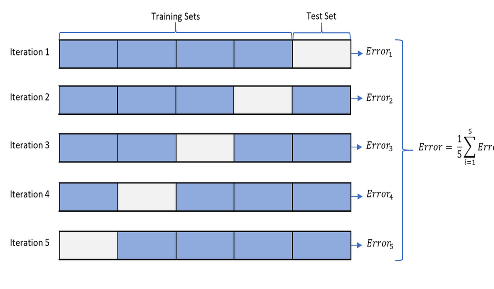
•Models exhibited high variance with different test sets.
•Cross-validation was employed to compare the general performance between models
•Cross validation ensured models were not overfit and could generalize well

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

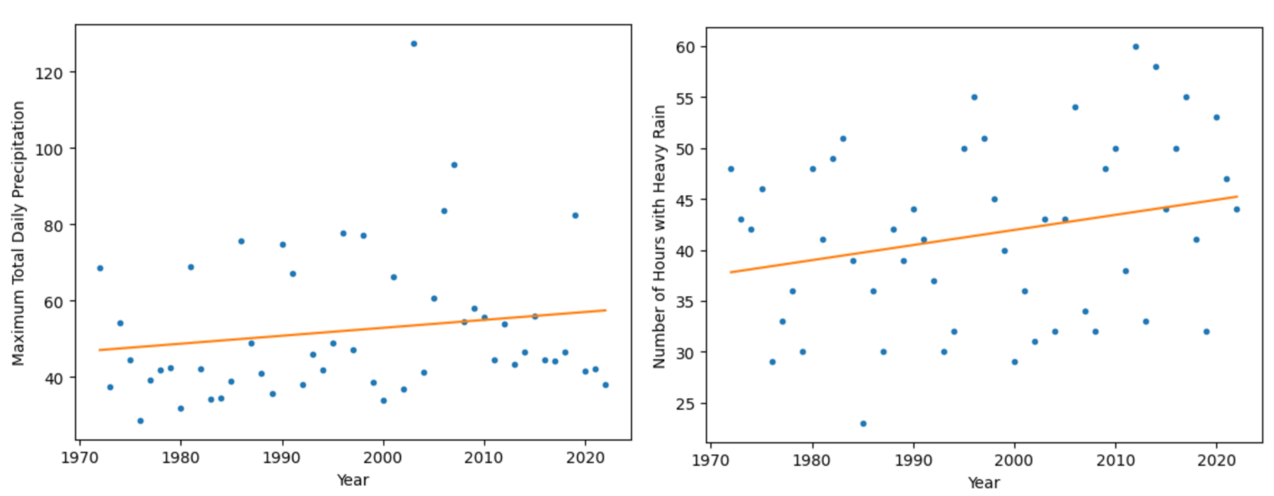
Random Seed 43 Results

	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

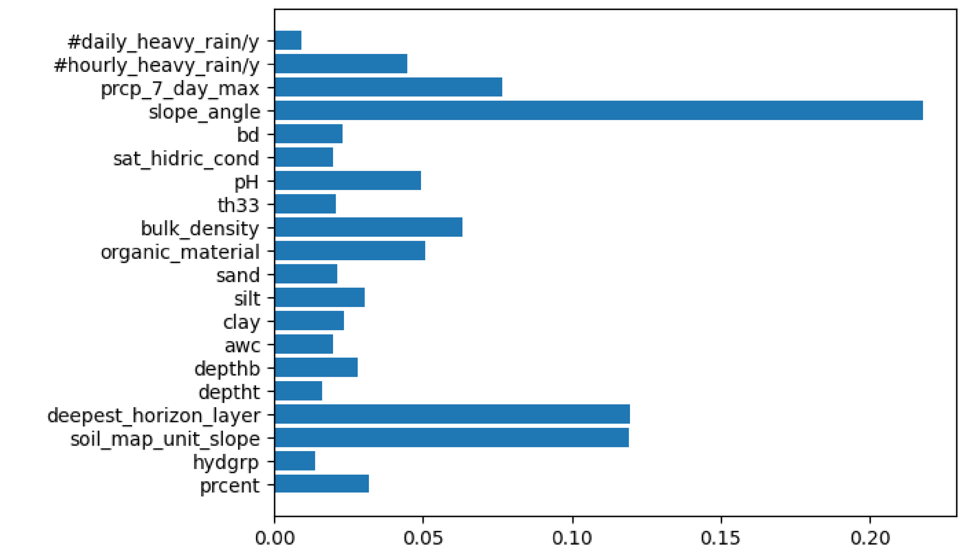
- To predict the maximum daily rainfall per year and the number of hours with heavy rain per year, we collect the data for these parameters in each location from 1972 to 2022
- We created a line of best fit through which we apply m*(2073)+b to get our prediction for the parameters in 2073



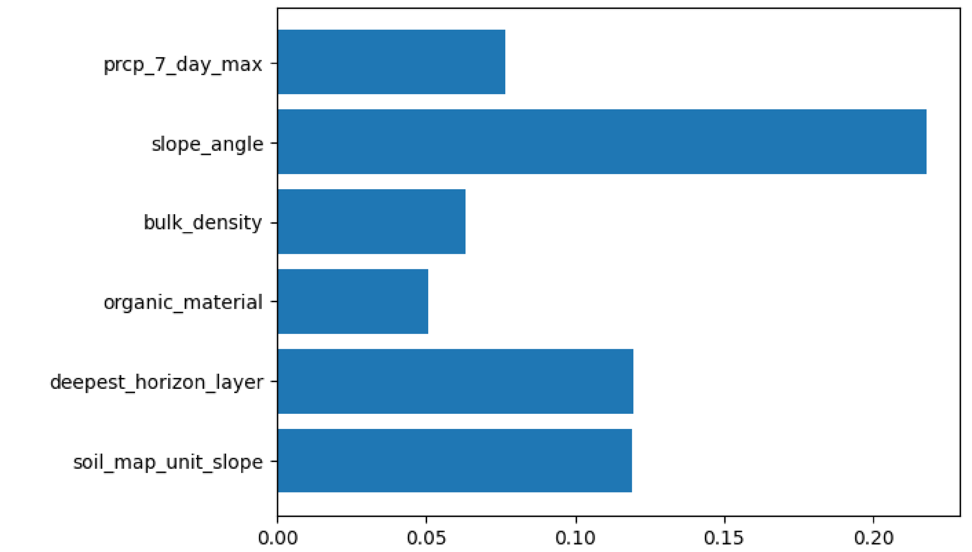
Feature Selection

Feature Engineering

- Feature engineering was used to choose the optimal number of features for model performance.
- The model with the top 15 features performed the best, but we will use the top 6 features for better interpretability.



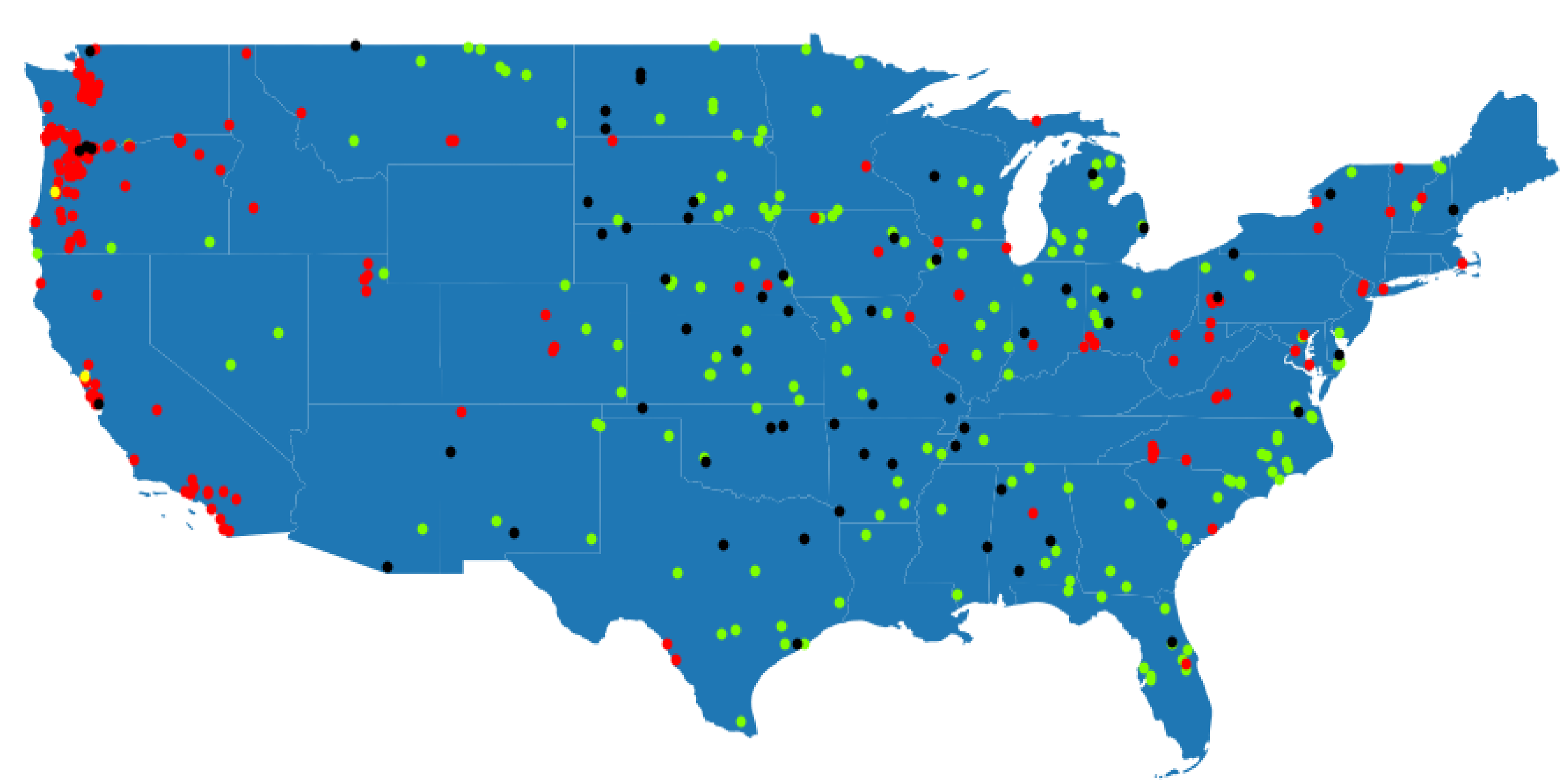
Random Forest Full Model
CV Accuracy: 0.8981



Random Forest Most Valuable Features
CV Accuracy: 0.9055

2073 Landslide Susceptibility Forecasting

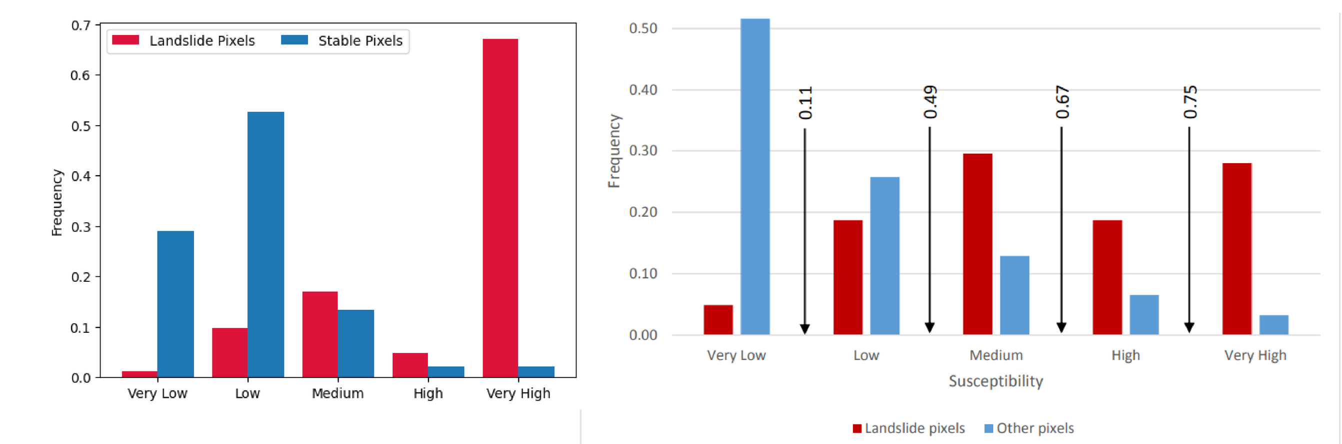
Random Forest prediction of Landslide Susceptibility in 2073 using predicted values for Maximum Daily Rainfall for each location in 2073 and the number of hours with heavy rain for each location in 2073



Results

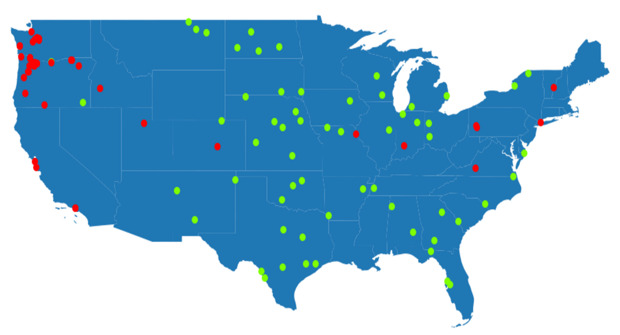
Comparison of our Model to NASA LHASA 1.1

- Due to underreporting, stability of pixels outside the landslide inventory is uncertain.
- Because of this, LHASA 1.1 chooses to predict the susceptibility of landslide pixels as opposed to other pixels instead of using an ROC curve 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 high or very high susceptibility compared to LHASA's nearly **50** percent.

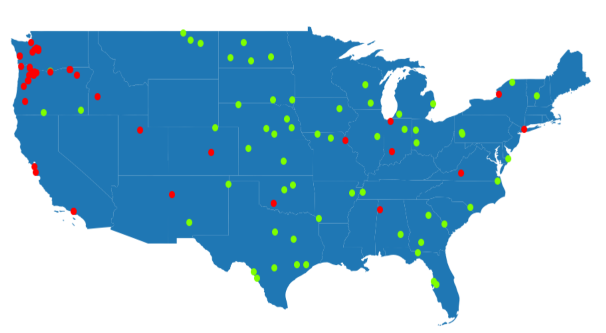


Test Set Performance

While the performance differs depending on the choice of test set, the mean test accuracy is approximately **90** 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

Conclusions & Future Work

•Limited number of samples prevented deep neural networks from converging and 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.
•We will verify input data collection methods against post-landslide field data.
•More parameters (e.g., soil moisture) and other triggers (e.g., earthquakes) can be considered to improve predictions.

References

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