

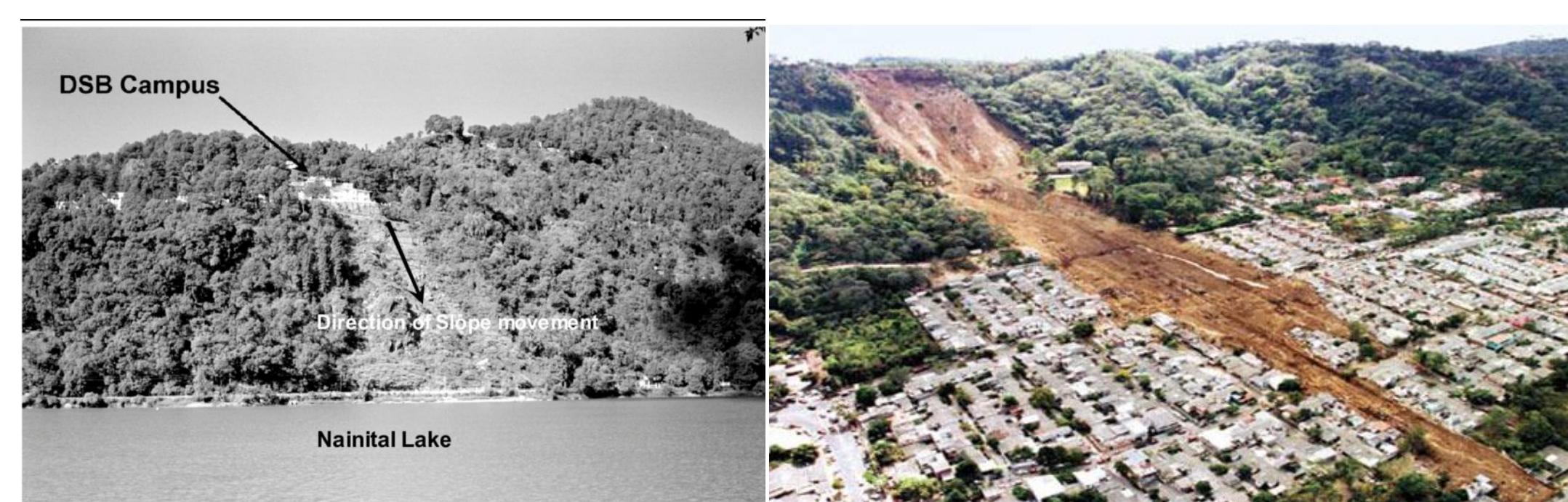
Applying Mathematical Modeling to Geophysical & Meteorological Data for Landslide Analytics & Forecasting

SciTeer Project #171603

Introduction and Background

Landslides are unpredictable natural disasters caused by changes in slope stability. The main causes of landslides are rainfall, human activity, and erosion.

- **18,000 deaths** and **4.8 million people** affected in past 2 decades
- **\$3.5 billion** in annual damages in the U.S.



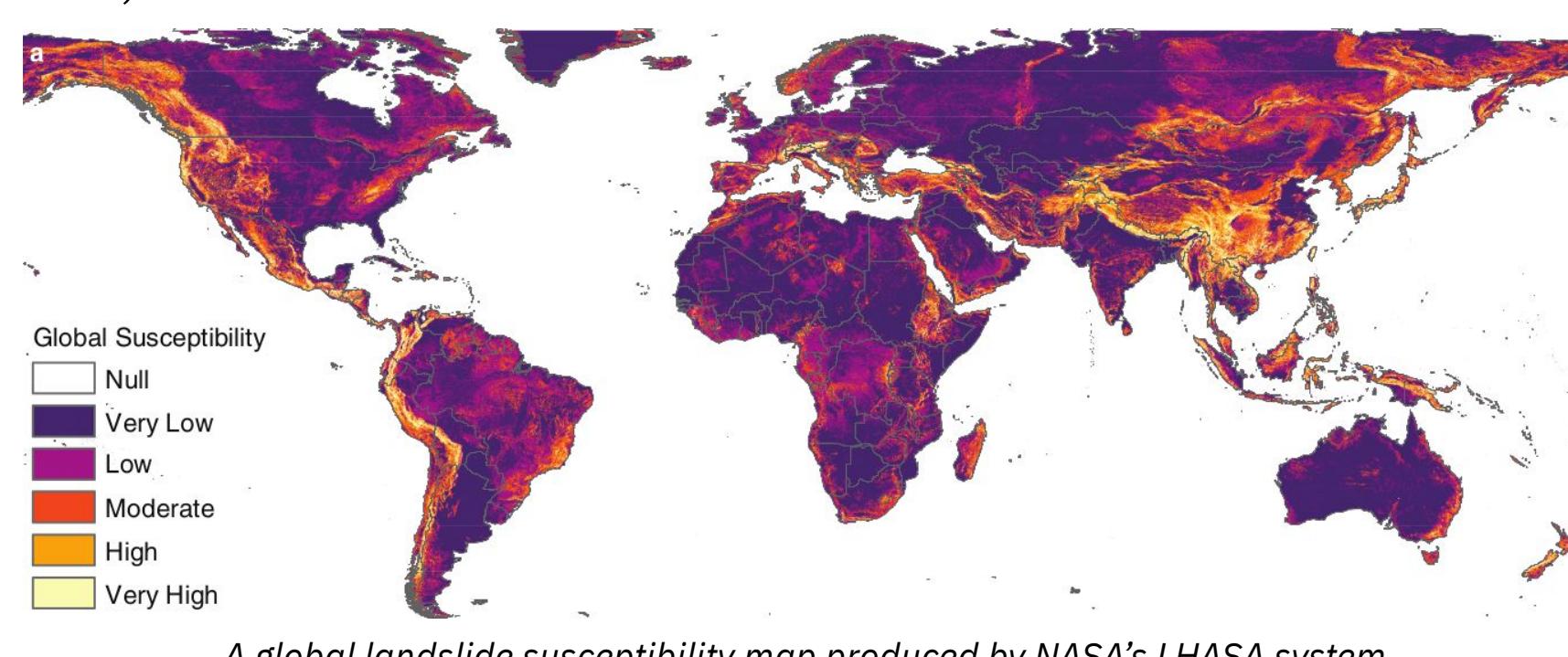
A photograph of a landslide near the DSB college campus in Nainital, India

A photograph of a landslide destroying homes and other infrastructure in its path.

Related Work

Landslide forecasting is a surprisingly difficult task due to lack of an obvious build-up and limited landslide-related data.

The only global-scale system for landslide forecasting that uses easily acquirable data is NASA's **LHASA** (Landslide Hazard Assessment for Situational Awareness).



A global landslide susceptibility map produced by NASA's LHASA system [Courtesy of LHASA]

LHASA uses a decision tree for generating landslide hazard "nowcasts". The ARI is calculated using IMERG data and a global susceptibility map is generated. These are used together to issue "nowcasts" of varying importance.

	1-day	3-day	7-day	FPR (%)
27	39	47	1	
24	35	40	1	
10	14	18	0.2	
8	14	16	0.2	

LHASA's evaluation statistics (above) and a snippet from the LHASA paper (below)

[Courtesy of LHASA]

issue and generate their own action plans. This system is not intended for local planning or to inform detailed infrastructure projects due to its geographic scope and spatial resolution. **LHASA is also not meant to be used as a warning or forecasting system**. This is due to the model latency (4-5 h from

Engineering Goal

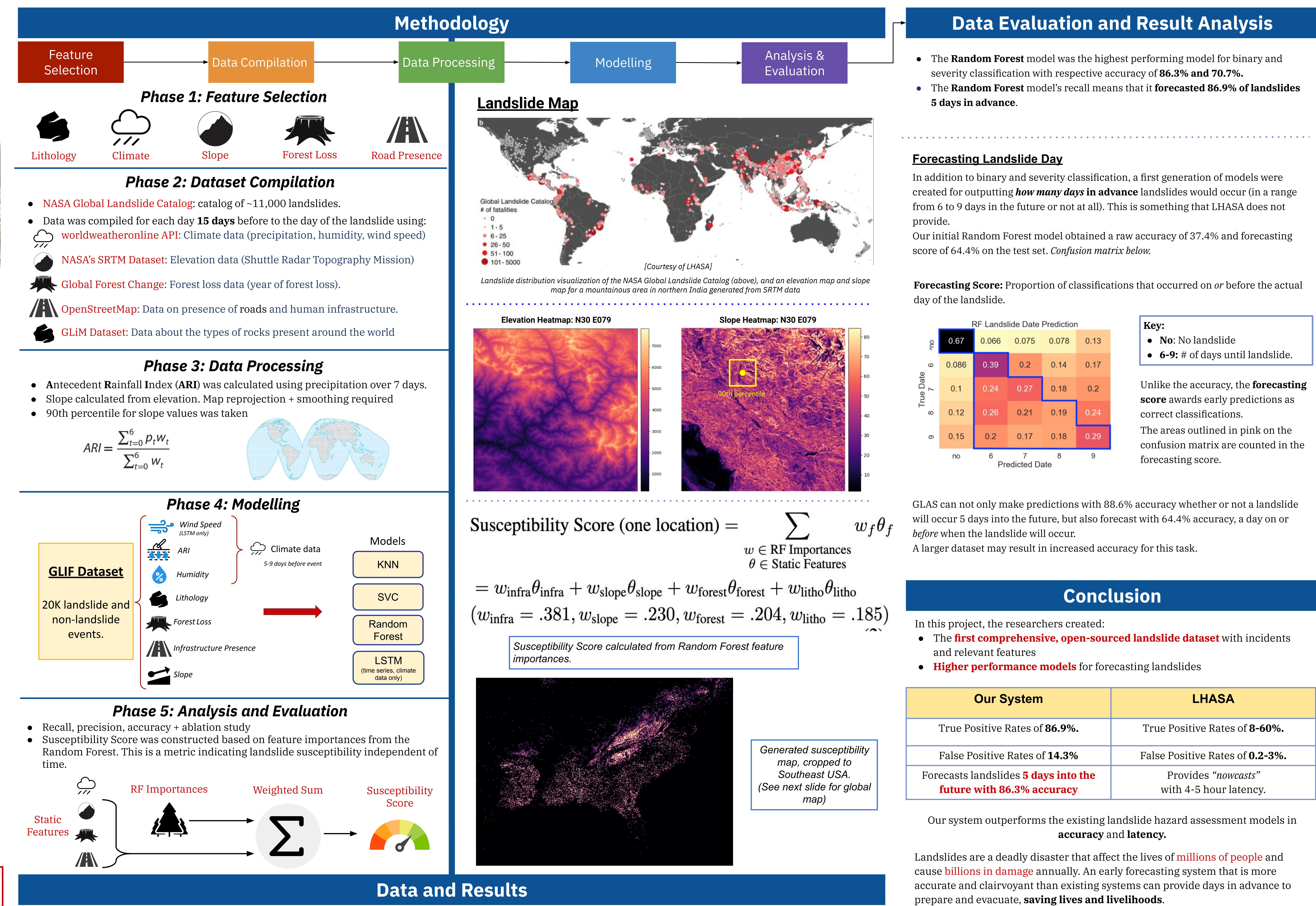
Engineering Goal:

This research's goal is to create **GLAS: A Global Landslide Analytics System** for:

- Landslide forecasting (can a landslide occur in the next 5 days?)
- Landslide severity analysis
- Terrain susceptibility analysis

Contributions:

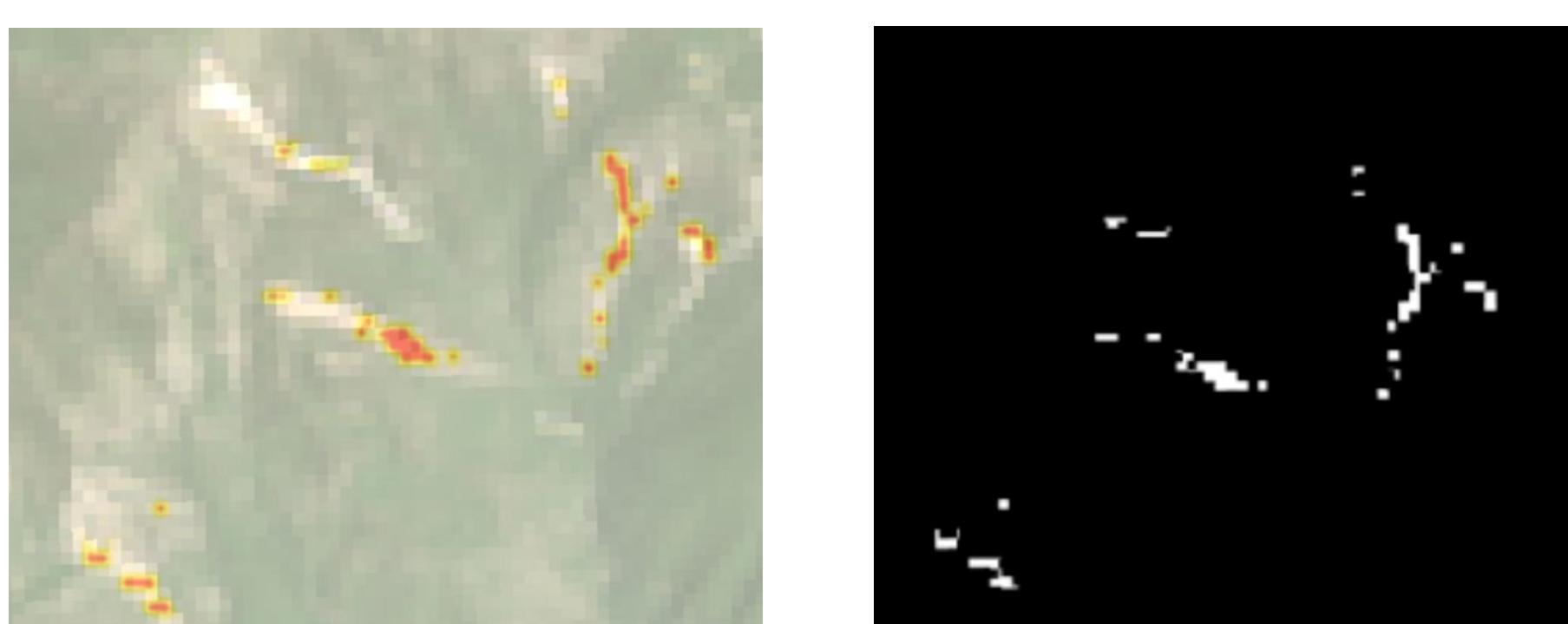
- First ever publicly available dataset of landslide incidents + relevant features (slope, weather, and terrain data)
- Proven mathematical algorithm for determining unreported past landslides
- More accurate, lower-latency landslide forecasting system



Data and Results

In addition to the time-independent, terrain-based susceptibility map, we designed a separate algorithm to determine landslide occurrence likelihood based on historical soil moisture and slope data. This was run through times around past major storms to find past unreported landslides to append to our GLIF dataset

The equation is on Slide 2.



(above left) A map for a region in Nepal depicting our algorithm's confidence heatmap.

(above right) A binary mask of the predicted landslide area. We used this to threshold a minimum area for a prediction to be considered a landslide.

Data Evaluation and Result Analysis

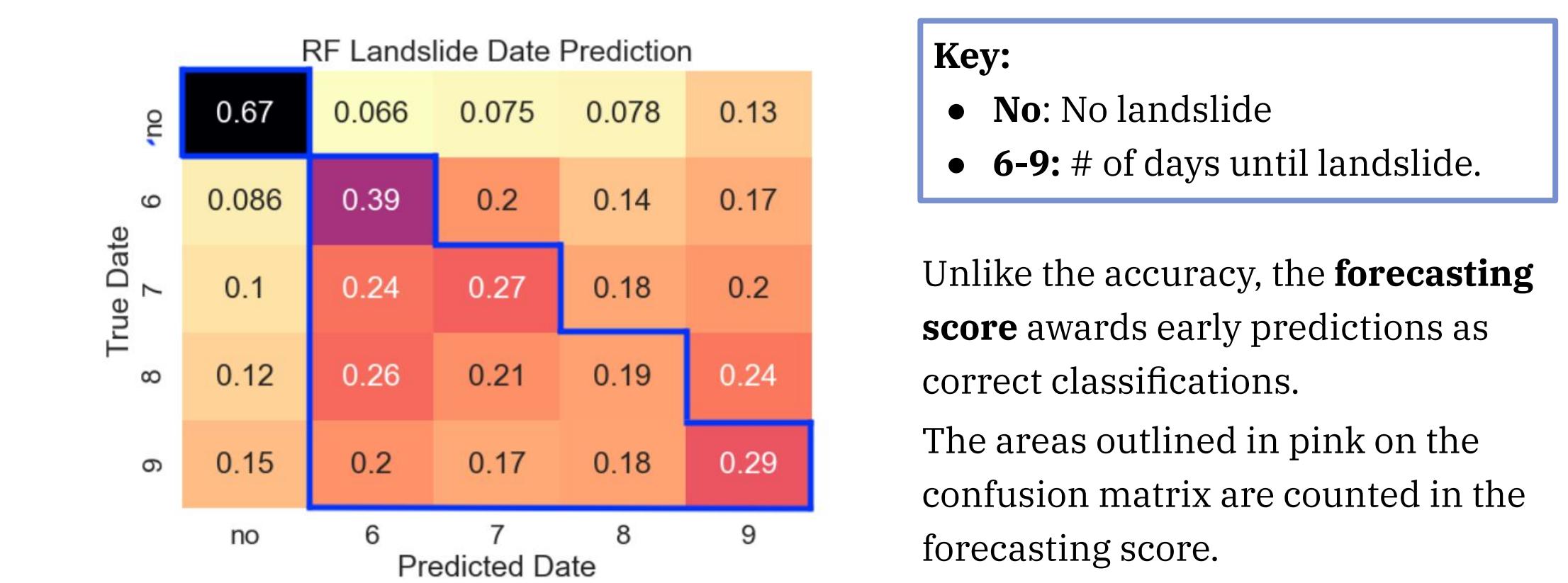
- The **Random Forest** model was the highest performing model for binary and severity classification with respective accuracy of **86.3%** and **70.7%**.
- The **Random Forest** model's recall means that it **forecasted 86.9% of landslides 5 days in advance**.

Forecasting Landslide Day

In addition to binary and severity classification, a first generation of models were created for outputting **how many days in advance** landslides would occur (in a range from 6 to 9 days in the future or not at all). This is something that LHASA does not provide.

Our initial Random Forest model obtained a raw accuracy of 37.4% and forecasting score of 64.4% on the test set. *Confusion matrix below*.

Forecasting Score: Proportion of classifications that occurred on *or* before the actual day of the landslide.



GLAS can not only make predictions with 88.6% accuracy whether or not a landslide will occur 5 days into the future, but also forecast with 64.4% accuracy, a day on or *before* when the landslide will occur. A larger dataset may result in increased accuracy for this task.

Conclusion

In this project, the researchers created:

- The **first comprehensive, open-sourced landslide dataset** with incidents and relevant features
- **Higher performance models** for forecasting landslides

Our System	LHASA
True Positive Rates of 86.9% .	True Positive Rates of 8-60% .
False Positive Rates of 14.3%	False Positive Rates of 0.2-3% .
Forecasts landslides 5 days into the future with 86.3% accuracy	Provides "nowcasts" with 4-5 hour latency.

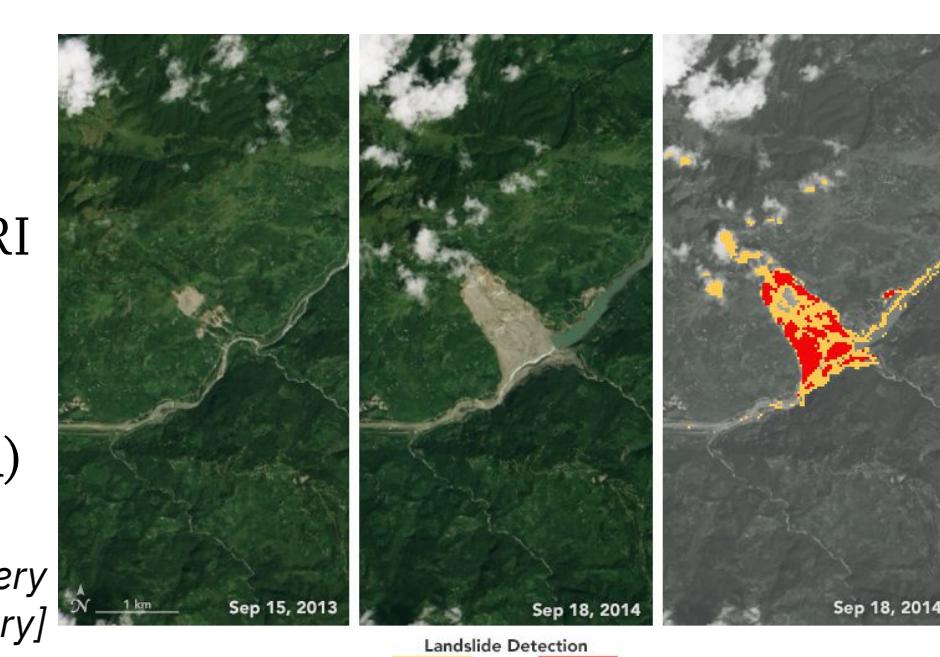
Our system outperforms the existing landslide hazard assessment models in **accuracy** and **latency**.

Landslides are a deadly disaster that affect the lives of **millions of people** and cause **billions in damage** annually. An early forecasting system that is more accurate and clairvoyant than existing systems can provide days in advance to prepare and evacuate, **saving lives and livelihoods**.

Future Work

In the future we'd like to incorporate more data sources and combine them in insightful ways. Some of the topics for future work include:

- Additional Features
 - Fault lines
 - Combined feature from Lithology + ARI
 - Satellite Images + Computer Vision
- Landslide web dashboard
- Different modeling techniques (lower FPR)



Landslide Detection using computer vision from satellite imagery [Courtesy of NASA Earth Observatory]

Landslide Detection Possible Probable

Random Forest Accuracies on Subsets

In addition to the whole GLIF dataset, two subsets of GLIF were created for the Random Forest model to train on to determine the model's performance with when non-landslide incidents were from random locations it hadn't seen before vs non-landslide incidents from former/future landslide locations.

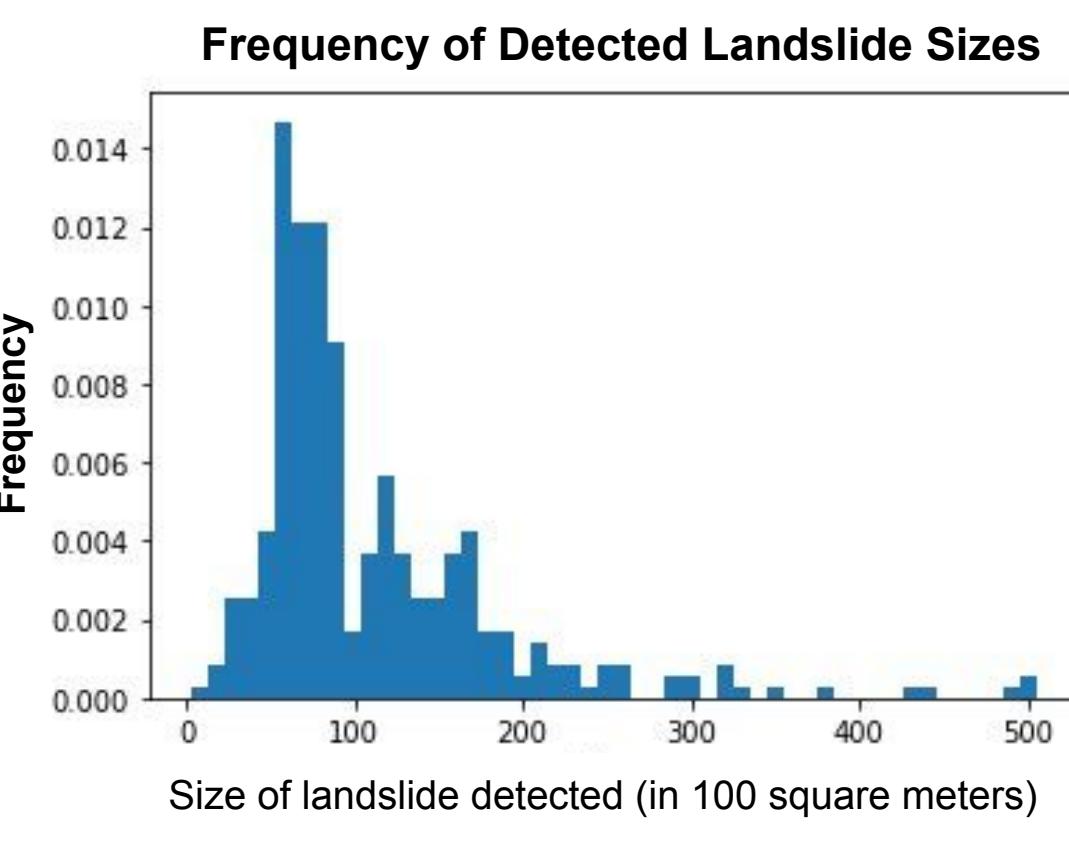
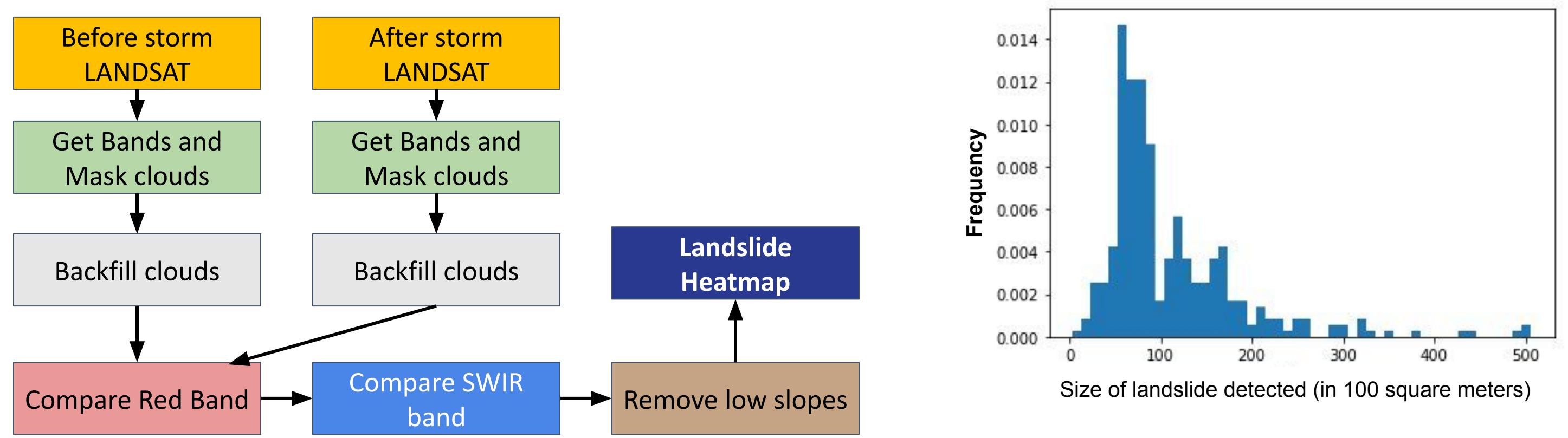
- Subset A:** 5,000 landslide events and 5,000 non-landslide events from random locations and times.
- Subset B:** 5,000 landslide and 5,000 non-landslide events from the same location, but at different times

	Binary Classification Accuracy	Severity Classification Accuracy	Date Prediction Early Forecasting Score
Entire GLIF Dataset	86.3%	72.6%	64.4%
Subset A	90.4%	71.8%	69.0%
Subset B	87.5%	67.5%	58.3%

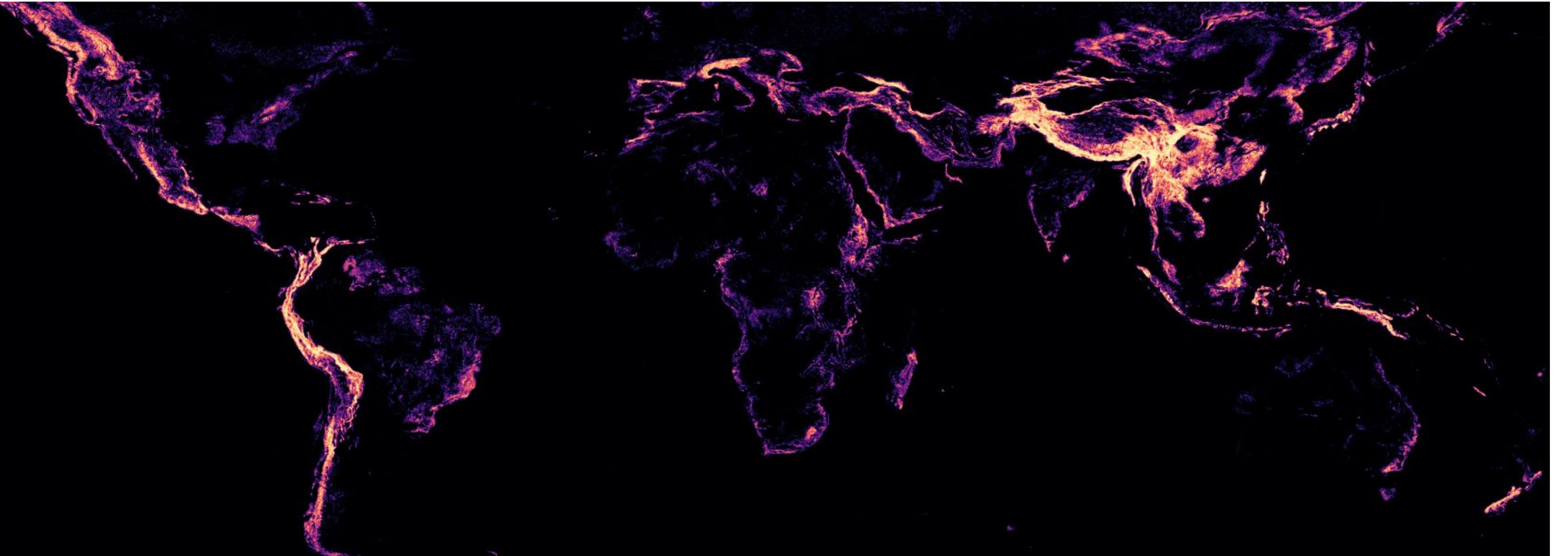
Algorithm for Mathematical Model Detecting Unreported Landslides using Historical Satellite Soil Moisture and Slope data

Algorithm Key Principles

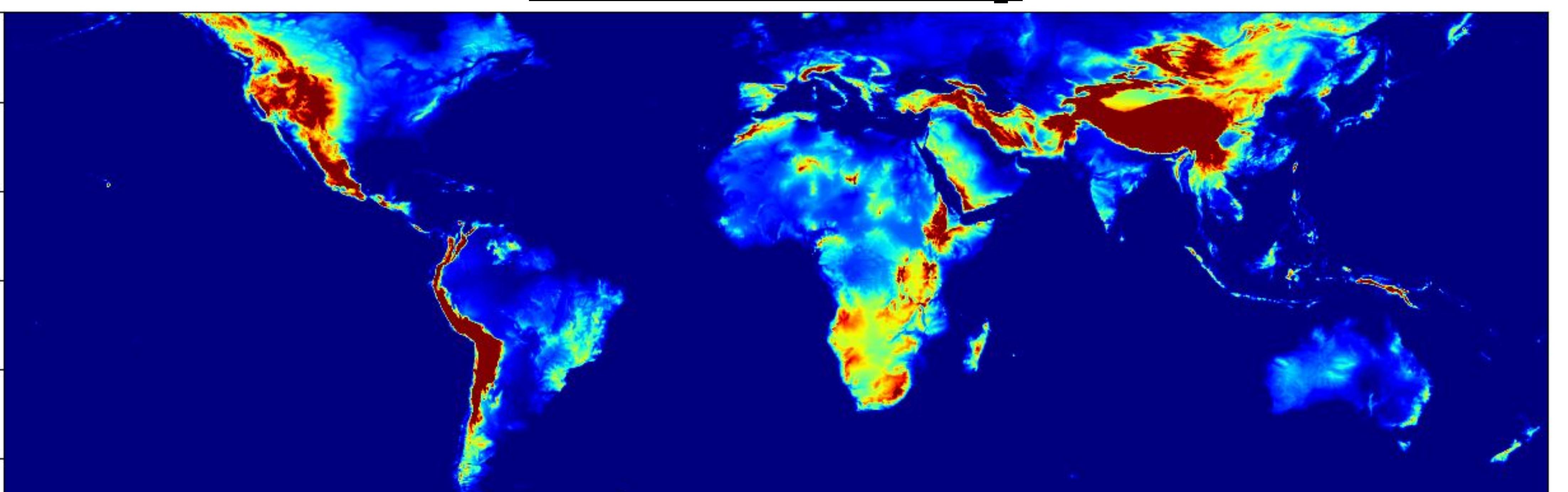
- Reflectance in Red Band (655nm) indicates bare earth exposure
- Changes in the SWIR bands 5 and 7 (860 and 2200 nm), indicate changes in soil moisture
- small slope values are not likely to be landslides



Our Generated Global Susceptibility Map



Global Susceptibility Map



Part of Dataset

id	date	lat	lon	country	fatalities	injuries	precip0	temp0	air0	humidity0	wind0	landslide	forest	ARI1	ARI0	slope	osm
														1	1	1	1
1069	9/4/16	46.726906	13.787332	Austria	0	0	3.8	67	1019	96	6	1	1	0.24081206	2.62486829	22.012	433
1855	3/23/17	49.726406	-116.91183	Canada	0	0	0.1	33	1020	100	8	1	1	0.54643713	0.25723617	34.101	445
797	10/20/09	18.5347	-72.4097	Haiti	4	0	3.4	87	1013	96	13	1	1	4.71700196	3.62581061	25.38	672
12967	12/31/09	4.42142905	-75.220709	Colombia	0	0	0	84	1017	96	11	1	1	0.3105131	0.17407389	29.537	681
13089	5/1/14	39.2902	-76.6651	United_States	0	0	0.4	77	1010	99	18	1	1	9.22206381	2.6665098	0	3909
9036	7/12/13	36.2629	-115.6158	United_States	0	0	0.2	84	1013	52	12	1	1	0.52317938	0.28614712	4.063	3930
7990	5/10/09	4.44059512	-75.243905	Colombia	0	0	0.8	82	1015	82	11	1	1	0.72870658	0.7786029	30.007	4597
3954	10/25/10	15.5227	-85.265	Honduras	0	0	4.7	86	1012	100	8	1	1	3.67732159	4.09146546	12.026	4597
2288	3/29/14	42.3946	-122.2137	United_States	0	0	2.1	47	1015	99	19	1	1	2.28104177	2.00814981	13.668	4611
13419	12/2/12	36.9933	-122.0206	United_States	0	0	9.5	60	1019	97	41	1	1	4.19111223	8.07512121	7.014	5236
5865	2/6/17	32.9134621	-107.7735	United_States	0	0	0	55	1016	52	32	1	1	0	0	0.853	5261
10892	9/15/16	20.8939	-156.6464	United_States	0	0	0.6	88	1016	77	25	1	1	1.69351435	1.01735069	0	5265
11940	3/21/12	43.6609	-123.3327	United_States	0	0	7.7	37	1015	100	11	1	1	5.8608551	6.69390133	22.175	5290
8746	8/1/13	44.4305	-118.1417	United_States	0	0	2.2	75	1014	74	16	1	1	0.33526948	1.53925477	11.203	5298
12993	4/17/14	43.4756	-110.7826	United_States	0	0	0	37	1022	98	10	1	1	0.437135	0.18141525	9.15	5423
8110	3/30/11	40.1625	-123.7874	United_States	0	0	0	65	1028	97	24	1	1	1.59220117	0.84505983	1.595	5453
7964	7/13/15	38.499	-80.7205	United_States	0	0	3.6	84	1013	99	12	1	1	2.72345949	3.56428892	12.562	5463
12732	3/4/15	37.336	-83.1295	United_States	0	0	3.4	52	1019	99	13	1	1	0.49019915	2.48276968	21.632	5472
3365	3/22/17	48.736721	-122.3590	United_States	0	0	0.7	51	1016	92	17	1	1	0.17214387	0.51866912	23.914	5495
5092	12/9/15	45.88	-118.9662	United_States	0	0	2.6	57	1012	79	43	1	1	2.18294704	2.50415515	22.002	5499
4140	2/28/14	45.6721	-121.877	United_States	0	0	0	46	1006	97	19	1	1	1.04431399	0.32729628	4.727	5588
6425	2/9/16	54.8366	-2.7812	United_Kingdom	0	0	0.7	41	984	92	39	1	1	0.84299744	0.81607201	0	5603
10633	8/5/13	-41.1212	146.1066	Australia	0	0	5.5	55	1007	93	42	1	1	2.32868108	4.67034097	0	5610
4716	9/7/08	30.0075	31.2774	Egypt	31	46	0	103	1009	90	12	1	1	0	0	12.823	5865
6041	8/29/15	45.2252	-122.3394	United_States	0	0	4.2	69	1011	97	25	1	1	0.19843933	2.82776051	18.61	5874
9835	12/8/15	47.488	-122.3633	United_States	0	0	7.9	56	1011	99	27	1	1	1.48247044	6.02616722	25.606	6197
7193	1/18/12	44.8746	-123.9314	United_States	0	0	12.5	50	1012	99	45	1	1	1.26321335	8.60157342	24.796	6208
12958	1/27/12	44.405951	-75.249405	Colombia	0	0	0.3	82	1015	84	4	1	1	0.90172491	0.5571938	29.888	6477
9009	11/20/12	42.9173	-124.1018	United_States	0	0	14.2	53	1008	99	24	1	1	14.6064212	13.367329	21.375	6480
3533	5/1/16	6.8586	37.7542	Ethiopia	42	0	0.1	84	1020	92	4	1	1	0.21856713	0.19425006	12.46	6610
12356	12/																

Applying Mathematical Modeling to Geophysical & Meteorological Data for Landslide Analytics & Forecasting

Ishaan Javali and Shrey Joshi

PLANO EAST SENIOR HIGH SCHOOL, Plano, Texas, US

Landslides result in billions of dollars of damage annually in the U.S. and affected 4.8 million people over 2 decades. Landslide detection is a critical task for the protection of human life and livelihood in mountainous areas. Existing landslide warning systems are inefficient in predicting landslide occurrences, in terms of accuracy, latency, or applicability. NASA's LHASA system has a 4-5 hour latency for landslide prediction with an 8-60% Probability of Detection (POD) and uses a 5-feature, static susceptibility map along with precipitation data. Other research papers collected specific data for limited regions. We propose GLAS, a Global Landslide Analytics System featuring higher-performance, scalable landslide forecasting models along with the first publicly available dataset of landslide events and features. Landslide incidence reports for GLAS were collected from the Global Landslide Catalog. For each landslide, indicative features of landslides were collected over a 15-day period: elevation, climate, forest loss, and street presence data. These features were compiled into a first-of-its-kind, public global dataset for landslide and non-landslide events. KNN, SVC, and Random Forest algorithms and an LSTM neural network were trained on the dataset to forecast whether there would be a landslide 5 days in advance, yielding an accuracy of 86.3% and a detection rate of 86.9% on the test set. These results exceeded LHASA's POD of 60%, thus providing people days to prepare and evacuate. We are currently refining and testing a mathematical model for finding unreported landslides by parsing historical satellite imagery, providing additional training data for our forecasting models.

Category

Mathematics

Components	Current Research Project	Previous Research Project Year: 2020-2021
1. Title	Applying Mathematical Modeling to Geophysical & Meteorological Data for Landslide Analytics & Forecasting	GLAS: A Global Landslide Analytics System
2. Change in goal/purpose/objective	<p>Goal:</p> <ul style="list-style-type: none"> Explore more landslide indicators, additional data sources, and perform landslide date forecasting Introduce a susceptibility mapping approach with static terrain features Parse through historical satellite data to find unreported landslides to use as additional training data with less self-report bias 	<p>Goal:</p> <ul style="list-style-type: none"> Create an initial dataset of global landslide indicators Train an initial iteration of forecasting models for binary and severity classification models
3. Changes in methodology	Model robustness was demonstrated by analyzing the dataset in two subsets: one containing random non-landslide instances, the other with non-landslide instances from landslide areas (different times). Hyperparameter optimization was improved. A landslide analytics dashboard was created	Data on landslide indicators was collected with Python scripts for landslide and non-landslide events. Landslide forecasting was done by training SVC, KNN, Random Forest, and LSTM models.
4. Variables studied	A Lithology feature was added to the dataset. The focus is on precipitation and slope data and improving the performance of the forecasting models through additional training data found by parsing through historical satellite data to find past unreported landslides.	Elevation, infrastructure presence, forest loss, precipitation, humidity, wind speed, air pressure, temperature data was collected and compiled into the dataset.
5. Additional changes	<p>The dataset consists of more accurate climate data.</p> <p>The model for forecasting the landslide date was improved.</p> <p>A mathematical algorithm was developed based on historical soil moisture and slope data to find unreported landslides to append to our dataset.</p>	NA

Attached are:

Abstract and Research Plan/Project Summary, Year 2020-2021

I hereby certify that the above information is correct and that the current year Abstract & Certification and project display board properly reflect work done only in the current year.

Ishaan Javali and Shrey Joshi Ishaan Javali 11/01/2021

Student's Printed Name(s) Signature Date of Signature

Abstract and Research Plan/Project Summary, Year 2020-2021
Page 45 International Rules: Guidelines for Science and Engineering Fairs 2021-2022 societyforscience.org/ISEF

This stamp or embossed seal attests that his project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

