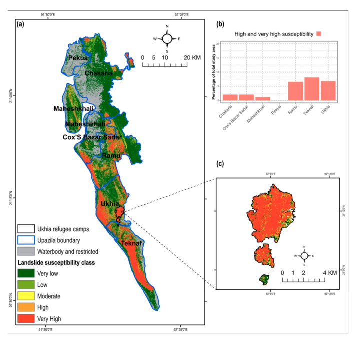


Machine Learning and Satellite Imagery to Landslide Susceptibility Prediction in Humanitarian Crisis

Md Salman Rahman
School of Mathematical and Statistical Sciences, University of Texas Rio Grande Valley
mdsalman.rahman01@utrgv.edu

Abstract

More than 980,000 stateless Rohingya refugees - an ethnic Muslim minority population from Rakhine State – have entered Bangladesh since August 2017 to flee Myanmar Army crimes against humanity. To make room for them, a large amount of deforestation and hill cutting took place in Bangladesh's Cox's Bazar District. The refugee camps are situated on slopes with unstable soil, making them particularly vulnerable to landslides caused by heavy rain. Landslides in the same region killed at least 160 people. This research aims to predict future landslide susceptibility and developing an early warning system for the Rohingya refugees and their host communities in Bangladesh. A novel method combining machine learning algorithm and satellite data, landslide inventory and susceptibility maps, rainfall thresholds, and a dynamic web-based alert system has been introduced to develop the landslide susceptibility map. We apply four machine learning algorithms: K Nearest Neighbor, Multi-Layer Perceptron (MLP), Random Forest (RF), and Support Vector Machine (SVM) to attain our objective. Finally, a combined landslide susceptibility map is prepared, integrating all the machine learning models to improve prediction accuracy. The Government and corresponding authorities can use our landslide prediction map to take action and save the lives of millions of Rohingya people from the catastrophic landslide.

Methods

Study Area

This study addressed the Cox's Bazar district, which is located in the south eastern region of Bangladesh. The study area lies between latitude $20^0 53^0 46.7''$ N and $21^0 14^0 29.8''$ N, and longitude $92^0 02^0 08.2''$ E and $92^0 18^0 27.0''$ E.

Data

No.	Conditioning Factor	Spatial Resolution	Variable Type
1	Aspect	30 m	Continuous
2	Elevation	"	"
3	Curvature	"	"
4	Slope	"	"
5	Stream Power Index (SPI)	"	"
6	Distance to stream	"	"
7	Land cover	"	Discrete
8	Normalized difference vegetation index (NDVI)	"	Continuous
9	Geology	"	Discrete
10	Soil type	"	"
11	Soil texture	"	"
12	Distance to road	"	Continuous

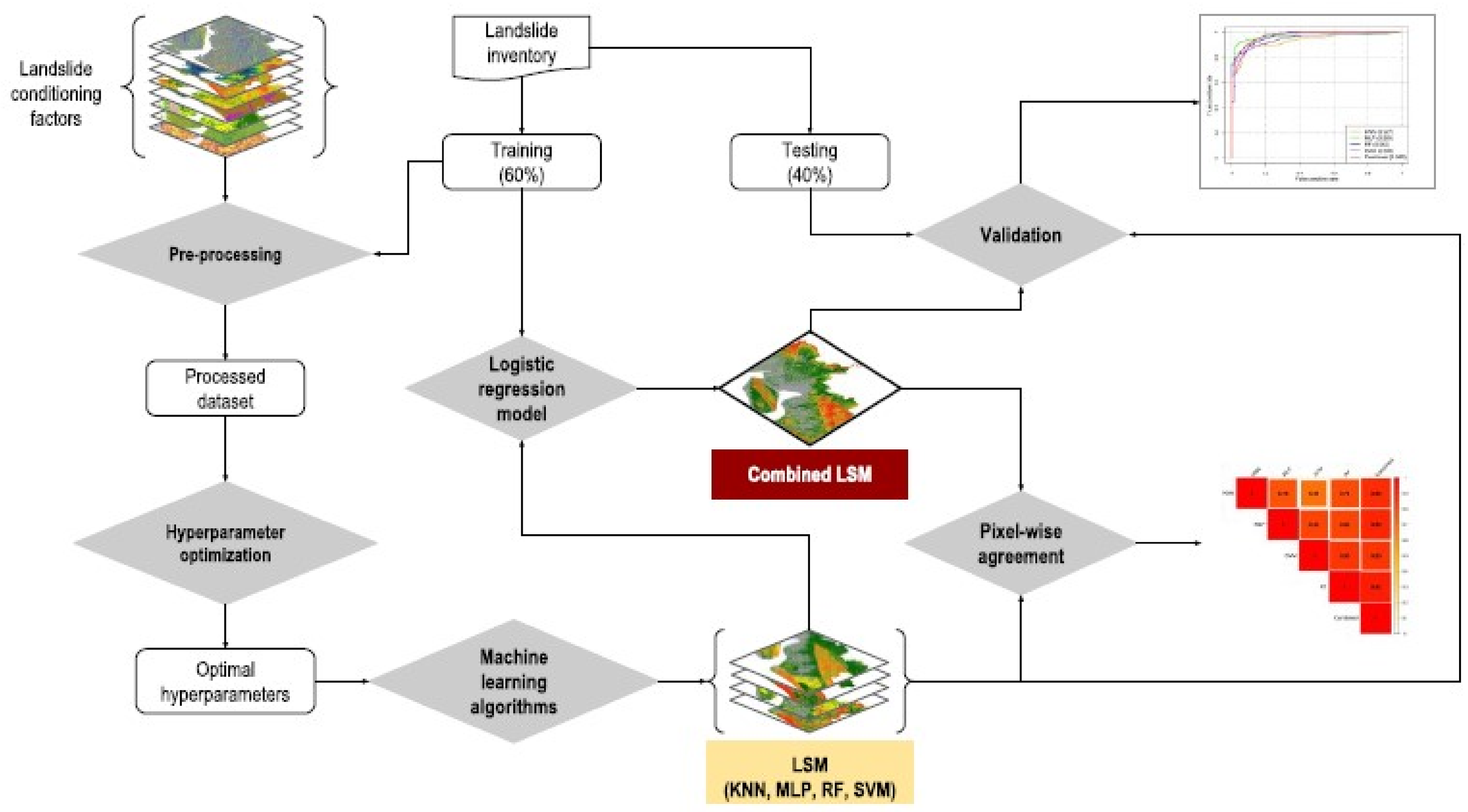
Models

The study was conducted in three stages. First, landslide susceptibility maps (LSMs) of the study area were developed using four machine learning algorithms: K-Nearest Neighbor (KNN), Multi-Layer Perceptron (MLP), Random Forest (RF), and Support Vector Machine (SVM). Second, the extent of spatial agreement of predicted patterns in LSMs was assessed by estimating the pixel-wise correlation of landslide probabilities obtained using various methods. Finally, an LSM was developed combining results from the four machine learning models.

Results and Discussion

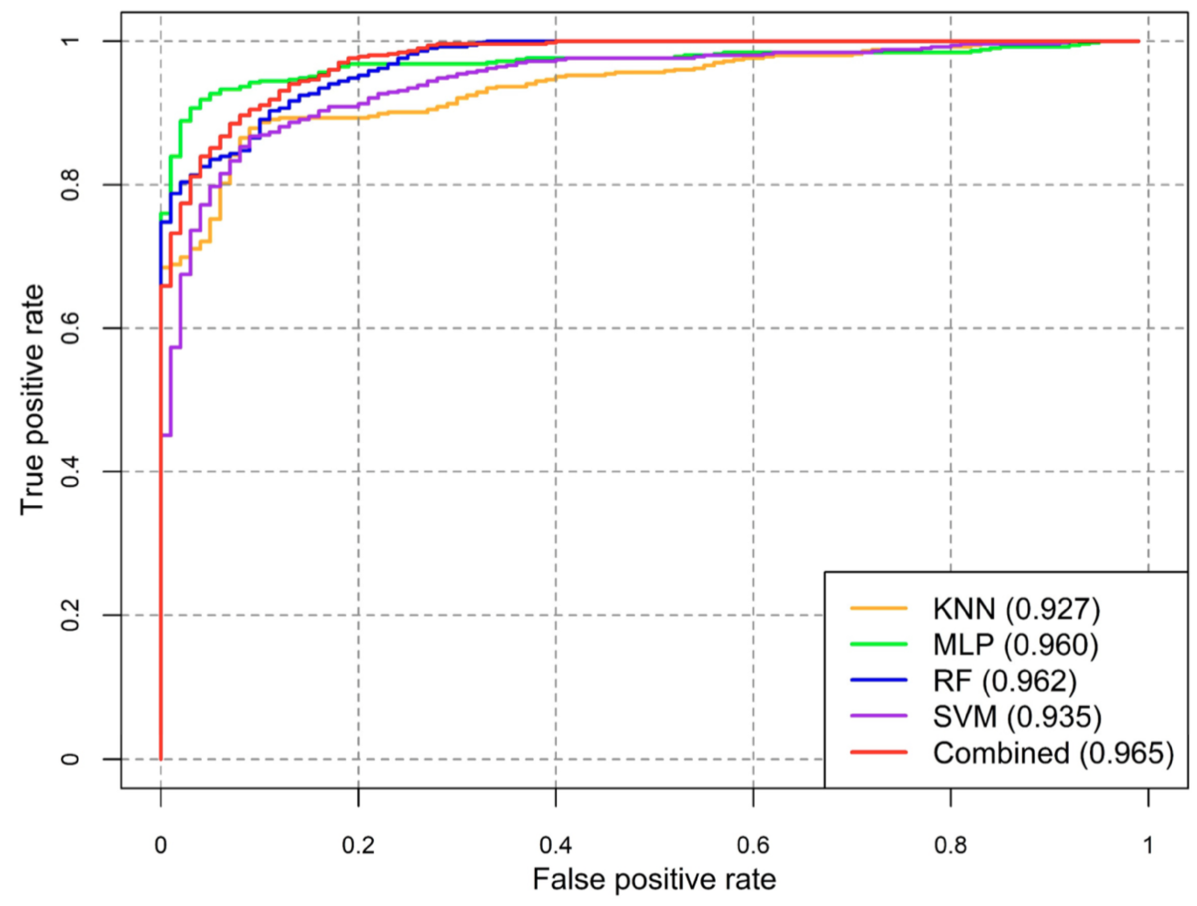
Landslide Prediction

The generated landslide probability maps were classified into five categories each by applying the Jenks natural breaks classification method in GIS: (i) very low (0–0.1), (ii) low (0.11–0.3), (iii) medium (0.31–0.5), (iv) high (0.51–0.85), and (v) very high (0.86–1). The proportion of landslide susceptible area varied from one model to another. Among all methods, the SVM resulted in the highest proportion of area (38.7%) susceptible to the landslide of 'high' and 'very high' severity, while the Random Forest (RF) algorithm yielded a relatively lower proportion (23.1%) of landslide susceptible area. Likewise, the ratio of the population exposed to 'high' and 'very high' landslide susceptible zones varied for different algorithms. For all the four methods, the percentage of landslide exposed population ranged between 34% to 48%.



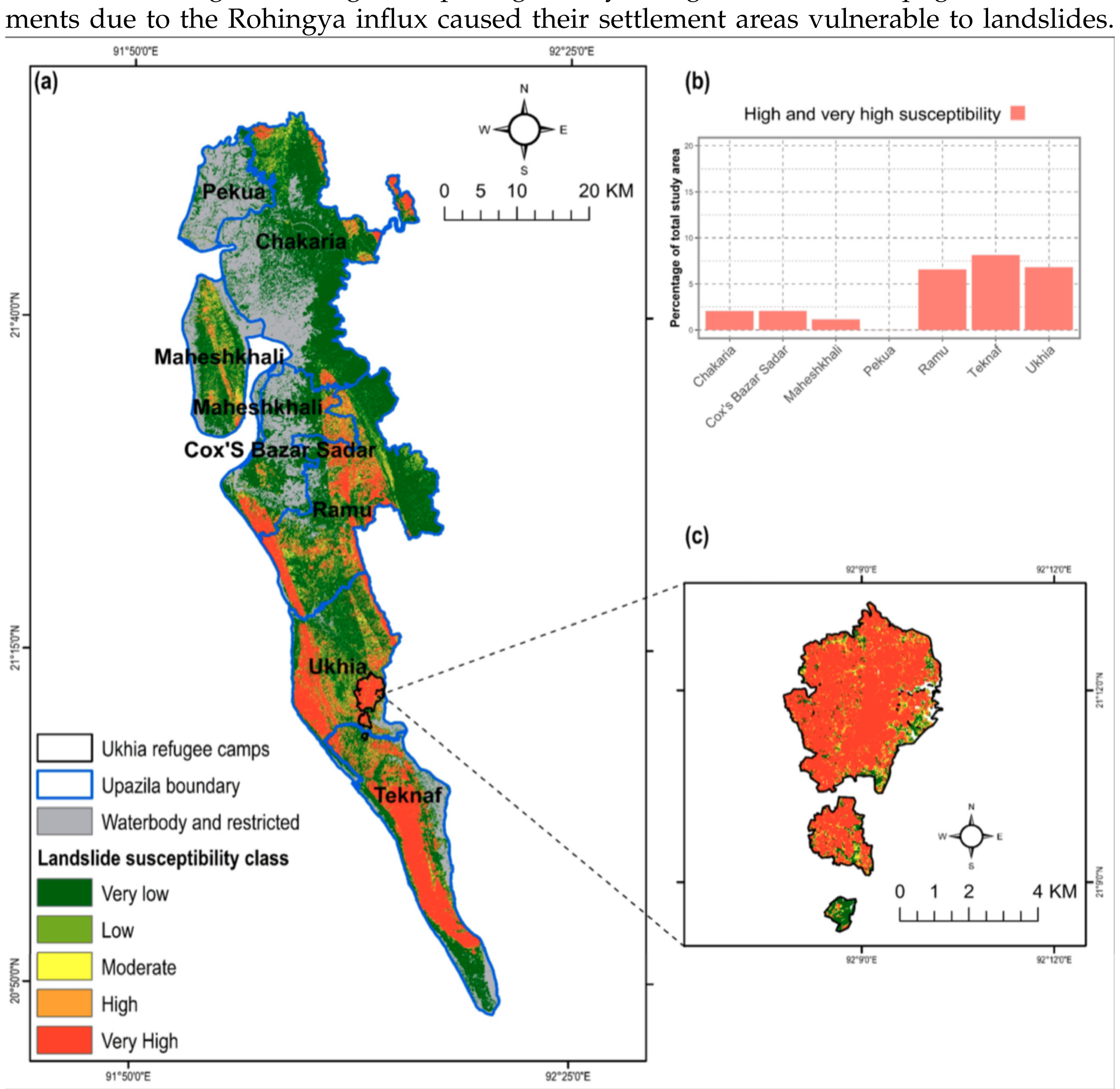
Models' Performance

Model	Overall Accuracy	Precision		F1-score		Recall	
		Non-Landslide	Landslide	Non-Landslide	Landslide	Non-Landslide	Landslide
KNN	0.9069	0.9227	0.9227	0.9015	0.9015	0.8811	0.8811
MLP	0.9545	0.9547	0.9547	0.9528	0.9528	0.9508	0.9508
RF	0.9663	0.9633	0.9633	0.9652	0.9652	0.9672	0.9672
SVM	0.9406	0.9385	0.9385	0.9385	0.9385	0.9385	0.9385



Landslide Susceptibility Mapping

Spatial heterogeneity in landslide prediction exists between the different machine learning-based approaches, an aggregated susceptibility map combining the outputs of all algorithms would minimize the uncertainty of individual methods. In this study, a regression-based approach was adopted. A multivariate logistic regression (LR) was established incorporating the binary landslide inventory data as the dependent variable and the results of the four landslide susceptibility models as independent variables. The outcome of the LR model is summarized in Model Performance. Among the four models, the MLP, RF, and SVM were statistically significant (p-value < 0.05). The coefficient of determinants (R^2) of 0.80 indicates a very good model performance. The extent of landslide susceptible areas varies in different sub-districts (Upazila) of Cox's Bazar. Teknaf Upazila is the most susceptible, where more than 8% of the total study area was susceptible to landslides of 'high and 'very high' severity. A substantial proportion of area (7% of the study area) in Ukhia sub-district was also susceptible. The Rohingya refugee camps in this area were located within high and very high landslide susceptible zones. Various recent studies also found that changes in the geomorphological, hydrological, and anthropogenic environments due to the Rohingya influx caused their settlement areas vulnerable to landslides.



Conclusion

This study is one of the pionner attempt to integrate results of multiple machine learning-based landslide susceptibility models to minimize uncertainties and improve landslide predictions. The modelling framework used in this study could be transferred to other landslide-susceptible regions. Landslide susceptibility maps can enable urban planners in identifying suitable areas for urban development. The combined landslide susceptibility map of Cox's Bazar district could be useful to policymakers and practitioners in sequencing and prioritizing interventions in managing landslides. The proposed model is an advancement in the existing landslide susceptibility models that intends to predict landslides more accurately. The results of this model could be utilized in improving the existing landslide early warning system, to strengthen landslide disaster risk mitigation strategies to support for the resilient future of inhabitants of the study area. By increasing machine learning and geoinformation in natural hazard assessment, environmental planning, and catastrophe risk reduction, the effort contributes to the UN Sustainable Development Goals.