Machine Learning-Based Prediction System for Rainfall-induced Landslides in Benguet First Engineering District

Zanya Reubenne D. Omadlao (Student 1), Nica Magdalena A. Tuguinay (Student 2),
and Ricarido M. Saturay, Jr. (Research adviser and teacher)

Philippine Science High School – Cordillera Administrative Region Campus, Department of Science and Technology, Purok 12, Irisan, Baguio City, 2600, Philippines

Corresponding Author Email: ric.psycarc@gmail.com

Abstract

A machine learning-based prediction system for rainfall-induced landslides in Benguet First Engineering District is proposed to address the landslide risk due to the climate and topography of Benguet province. It is intended to improve the decision support system for road management with regards to landslides, as implemented by the Department of Public Works and Highways Benguet First District Engineering Office. Supervised classification was applied to daily rainfall and landslide data for the Benguet First Engineering District covering the years 2014 to 2018 using scikit-learn. Various forms of cumulative rainfall values were used to predict landslide occurrence for a given day. Following typical machine learning workflows, rainfall-landslide data set was divided into training and testing data sets. Machine learning algorithms such as K-Nearest Neighbors, Gaussian Naïve Bayes, Support Vector Machine, Logistic Regression, Random Forest, Decision Tree, and AdaBoost were trained using the training data sets, and the trained models were used to make predictions based on the testing data sets. Predictive performance of the models vis-a-vis the testing data sets were compared using true positive rates, false positive rates, and the area under the Receiver Operating Characteristic Curve. Predictive performance of these models were then compared to 1-day cumulative rainfall thresholds commonly used for landslide predictions. Among the machine learning models evaluated, Gaussian Naïve Bayes has the best performance, with mean false positive rate, true positive rate and area under the curve scores of 7%, 76%, and 84% respectively. It also performs better than the 1-day cumulative rainfall thresholds. This research demonstrates the potential of machine learning for identifying temporal patterns in rainfall-induced landslides using minimal data input -- daily rainfall from a single synoptic station, and highway maintenance records. Such an approach may be tested and applied to similar problems in the field of disaster risk reduction and management.

Keywords: Benguet; landslide prediction; machine learning; Philippines; Python; scikit-learn

Introduction

The Philippines' Department of Public Works and Highways (DPWH) is primarily responsible for the management of national highways. Landslides are common in Benguet Province due to the climate and topography of the area, and many are rainfall-induced (Nolasco-Javier et al., 2014). These hazards strongly affect highway operations and maintenance, and pose serious risks to road users and nearby residents. While landslide susceptible areas have been mapped by the Mines and Geosciences Bureau (Lands Geological Survey Division, 2015) and Project NOAH (Eco et al., 2015), temporal forecasts of landslides are mainly generalized warnings embedded in official weather updates during significant weather disturbances. A rainfall-based prediction system can

enhance the existing DPWH decision support system for road management in relation to landslides.

Machine learning (ML) models have seen increasing use in the area of time series forecasting (Bontempi et al., 2012). Rainfall data collected and compiled by the Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) is essentially a time series. However, it has only been used to a limited extent only in post-landslide event analyses (Nolasco-Javier et al., 2014; Nieva-Nishimori, 2018; and Agoot 2018), and has yet to be integrated in, and institutionalized for landslide forecasting.

The researchers proposed to develop a system for predicting rainfall-induced landslides using ML models, rainfall data from PAGASA, and landslide data derived from DPWH highway maintenance records. Specific objectives were to generate relevant features and labels for input to ML algorithms; to evaluate the predictive performance of the resulting ML models; and to compare ML models with previously proposed rainfall thresholds.

Practical access, consistency, and continuity of records constrained this research to the use of rainfall data from the PAGASA Baguio Synoptic Station, and maintenance records from the highways within the Benguet First Engineering District (BFED) from the years 2014 to 2018. Results are therefore primarily applicable to national highways managed by the DPWH-BFED Office, and must not be haphazardly applied to other areas not covered by this study.

Methodology

Acquisition of data

Highway maintenance records were obtained from the DPWH-BFED Office. The national roads managed by the Office are found in the southern portion of Benguet Province. Daily records of maintenance were inspected to evaluate whether a landslide had occurred or not. All dates with reported soil collapse, landslide, roadslip, mud flow, debris flow, riprap collapse, rockfall, slope collapse, land/rock slide, and/or land/debris flow were classified by the researchers as landslide events. The rest of the dates with no such remarks were classified as non-landslide events. The data was digitized, containing a date column, and a boolean column, wherein landslide and non-landslide events are represented by 1 and 0, respectively. For this

study, landslide events do not indicate the number of landslide events for a given date.

Rainfall data were obtained from the PAGASA-Baguio Synoptic Station, and digitized. It consists of daily rainfall measurements made every 8:00AM representing accumulated rainfall in the last 24 hours.

Rainfall and landslide data were combined, with columns for date, 24-hour rainfall, and landslide event. Data covers the years 2014 to 2018 (Fig. 1).
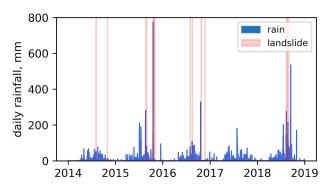

Figure 1. Daily rainfall and landslide data from 2014 to 2018 for the national highways within the Benguet First Engineering District. For this period, there are 34 days classified as landslide events (red bars). Maximum daily rainfall is 754 mm. Data derived from PAGASA Baguio Synoptic Station and DPWH-BFED Office.

Supervised classification using scikit-learn

This study approached landslide prediction as a machine learning problem involving supervised classification, wherein data (rainfall) comes with additional attributes or labels (landslide or no landslide) which need to be predicted by the model (scikit-learn developers, 2019). The main implementation for the solution uses scikit-learn, a Python-based module for machine learning (Pedregosa et al., 2011), and is supported by pandas, a Python library of data structures and statistical tools (McKinney, 2010). An overview of the machine learning implementation is shown in Figure 2

.
Combinations of cumulative and offset rainfall, similar to the recent- and antecedent-cumulative rainfall thresholds of Scheevel, Baum, Mirus, & Smith (2017), were used as features or predictors. From the PAGASA daily rainfall records (Fig. 1), cumulative-offset rainfalls, as a function of number of days for accumulation (1, 2, 5, 10, 15, 30 days), and number of days offset from the present record (0, 1, 2, 5 days), were generated and combined,

corresponding to 24 features. These features and the corresponding labels (landslide vs no landslide) were then combined in a single data set for use in the succeeding ML operations. Each feature was assigned to a single column in the data set. The final data set for analysis consists of 1757 rows corresponding to days, and 26 columns for the date, the 24 features, and the label.
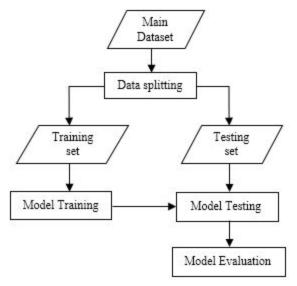

Figure 2. Overview of machine learning implementation.

Data splitting

It is common practice in machine learning to split the data into training and testing sets (scikit-learn developers, 2019) (Fig. 2). The former is used to teach or train the ML algorithms, while the latter is used to evaluate how well the ML algorithm has learned from the training dataset. For this study, a stratified random split was done to produce a training set from 75% of the original data set, and a testing set from the remaining 25%. A stratified split was used to ensure that proportional amounts of landslide events were incorporated in both training and testing sets, considering there are only 34 landslide events for the whole data set.

Model Training and Testing

The following ML algorithms: K-Nearest Neighbors (kNN), Gaussian Naïve Bayes (GNB), Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Decision Tree (DT) and AdaBoost (AdaB) (Pedregosa et al., 2011; Pham et al., 2016), were trained using the training set (Fig. 2).

Default parameters for each algorithm were used, and were not modified. The trained models were then used to predict landslide occurrence given the testing set (Fig. 2).

Model Evaluation

A confusion matrix was used to characterize the predictions from the testing set as true positive, *tp*, (actual landslide event correctly predicted), true negative, *tn*, (actual non-landslide event correctly predicted), false positive, *fp*, (false alarm) and false negative, *fn*, (missed prediction). From the matrix, true positive rates ($TPR=tp/(tp+fn)$) and false positive rates ($FPR=fp/(fp+tn)$) were computed. Good models should have high values of TPR and low values of FPR (Begueria 2006).

A Receiver Operating Characteristic (ROC) Curve was used to produce Area Under the Curve (AUC) scores from the TPR and FPR scores . Good models should have FPR vs TPR plotting close to the upper left corner of the ROC plot, and thus maximize the AUCs such that the score is close or equal to 1 (100% TPR, 0% FPR) (Fawcett 2005).

Split-Train-Test-Evaluate Iterations

For each ML algorithm, the sequence of splitting, training, testing, and evaluation (Fig. 2) was repeated 100 times to produce 100 unique training and testing sets, and distributions of FPRs, TPRs, and AUCs. Scores were plotted as stacked histograms and scatter plots. The means and standard deviations were also computed.

Statistical analysis using the scipy library (Oliphant, 2007) was done to characterize the differences among and between the resulting distributions. FPRs, TPRs, and AUCs for all models were subject to the non-parametric Kruskal-Wallis test (Kruskal & Wallis, 1952). For post-hoc comparisons between models, a T-test for non-uniform variance was used (Welch, 1947). Hypothesis tests were evaluated at 95% significance level.

Comparison with Previously Proposed Rainfall Thresholds

The original rainfall data set was also used to generate landslide predictions based on 1-day cumulative rainfall thresholds, similar to those proposed or mentioned by Nolasco-Javier et al. (2014), Agoot (2018), and Nieva-Nishimori (2018). A range of threshold values from the 0th to the 100th percentile of the 1-day cumulative

rainfall was evaluated, and the corresponding TPRs, FPRs, and AUCs were analyzed.

Results

True Positive Rates, False Positive Rates, and Areas Under the Curve

The stacked histograms (Figs. 3 and 4) show the distribution of FPR and TPR for the different models. These two sets of values are plotted against each other in an ROC plot (Fig. 5). A separate stacked histogram shows the distribution of the resulting AUCs, as a function of FPR and TPR (Fig. 6). Table 1 shows the means and standard deviations of the FPRs, TPRs, and AUCs of the evaluated models. For clarity of figures and brevity of the table, results for SVM and LR models, both of which had zero true and false positive rates, are not included.

Statistical Tests

Kruskal-Wallis tests for FPRs, TPRs, and AUCs all return p-values equal to zero, suggesting that at least one model's performance is different from those of the others. Post-hoc comparisons between models using T-tests show most groups have different mean FPRs, TPRs, and AUCs, except for FPRs of DT and AdaB (p-value=0.085); TPRs of DT and AdaB (p-value=0.284); and AUCs of DT and AdaB (p-value=0.304).

Comparison with Previously Proposed Rainfall Thresholds

The ROC plot in Figure 7 overlays the TPR-FPR pairs derived from predictions using only the 1-day cumulative rainfall. Rightmost and leftmost points correspond to the 0th percentile (0mm) and 100th percentile (754mm). The threshold value of 70mm (Nolasco-Javier et al., 2014) and greater (Agoot, 2018; and Nieva-Nishimori, 2018) are above the 97th percentile, and plot in the region of TPR<20% and near-zero FPR.

Discussion

Results show that the GNB model has the highest TPRs among the ML models (Fig. 4, Tab.1). It also has the highest AUCs, despite having the highest FPRs (Figs. 3, 6; Tab.1). Its scores are also significantly different from those of the other models. Assuming equal weights given to TPRs and FPRs, the GNB model, based on the AUCs, may
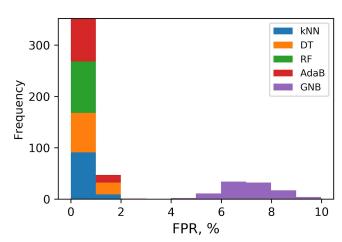

Figure 3. Stacked histogram of false positive rates (FPR) for the various machine learning algorithms evaluated.
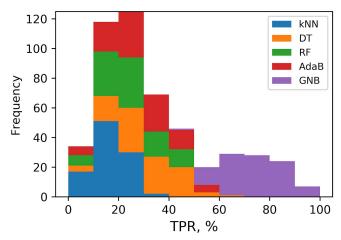

Figure 4. Stacked histogram of true positive rates (TPR) for the various machine learning algorithms evaluated.

be considered as the best model among those evaluated for predicting landslides in the study area.

The 1-day cumulative rainfall thresholds generally plot to the right of the ROC plots of the ML models, closer to the random guess line (Fig. 7). This indicates less predictive performance compared to the ML models in this study. It is notable that the 1-day cumulative rainfall as a landslide threshold, is often cited by authorities and the media, and yet its performance has yet to be systematically assessed.

This study also highlights the potential of machine learning in maximizing existing datasets already being collected by government agencies. Despite known concerns regarding the required spatial density of rainfall

observations to increase the reliability of landslide predictions (Nikolopoulos, Borga, Creutin, & Marra, 2015), the study demonstrates how data even from a single rain gauge station (Baguio Synoptic Station) was used to make good predictions at low data acquisition costs.
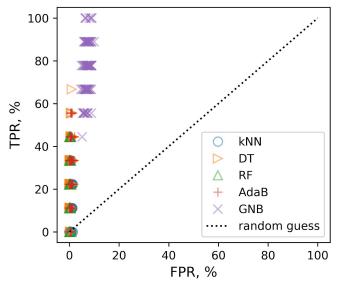

Figure 5. True and false positive rates, plotted in a Receiver Operating Characteristic (ROC) plot for computing the area under the curve (AUC).
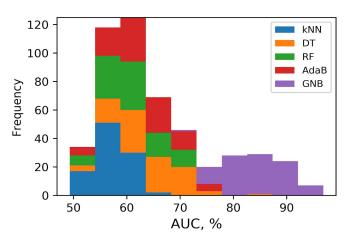

Figure 6. Stacked histogram of Areas Under the Curve (AUC) for the various machine learning algorithms evaluated.

It must be noted that the predictions in this study were made without further optimizing the data features and the machine learning algorithms. Thus prediction performances can still be further increased. The predictions in this study, however, are temporal in nature, meaning the

Table 1

Predictive Performance of Various Machine Learning Models

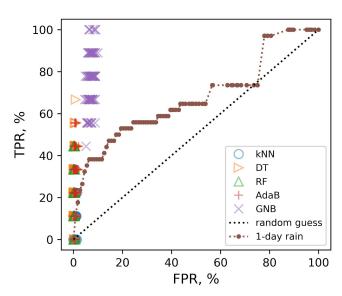| Model	FPR (%)		TPR (%)		AUC (%)	
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
GNB	7.2	1.4	75.9	12.9	84.4	6 .3
DT	0.7	0.4	28.7	14.5	64.0	7.2
AdaB	0.7	0.4	25.6	13.5	62.5	6.7
RF	0.1	0.2	21.8	12.7	60.8	6.4
kNN	0.5	0.4	13.0	8.1	56.3	4.0

Figure 7. True and false positive rates for predictions using only the 1-day cumulative rainfall, overlaid to data in Figure 5. This shows lower predictive performance of 1-day cumulative rainfall thresholds compared to machine learning models.

system predicts when a landslide is likely to occur in the study area, but it does not answer where it will happen. Hence, this system should be integrated with landslide susceptibility maps which show where landslides are likely to happen.

Summary and Conclusion

The study was able to set-up a machine learning system which uses rainfall from a single station, and road maintenance records to predict landslides for national highways within the Benguet First Engineering District. The Gaussian Naive-Bayes algorithm has the best predictive performance among the evaluated models, with mean TPR=76%, mean FPR=7%, and mean AUC=84%. This model also outperforms the traditionally used 1-day cumulative rainfall thresholds, for any value of the cumulative rainfall.

Further improvements in predictive performance can be achieved with further optimizations of data features and the machine learning algorithms. Temporal predictions from this system can be integrated with landslide susceptibility maps to provide a forecast for when and where landslides may occur in the area. Results are, however, primarily applicable to national highways managed by the DPWH-BFED Office, and must not be haphazardly applied to other areas not covered by this study.

This research demonstrates the potential of machine learning for characterizing temporal patterns in rainfall-induced landslides using minimal data input. This approach may be tested and used for other disaster risk reduction and management problems which require actual observational data.

Acknowledgements

References

Agoot, L. (2018, August 15). *Incessant heavy rains cause landslides, riprap collapse in Cordillera.* Retrieved May 15, 2019, from https://www.pna.gov.ph/articles/1044915

Begueria, S. (2006). Validation and evaluation of predictive models in hazard assessment and risk management. *Natural Hazards*, 37(3), 315–329. https://doi.org/10.1007/s11069-005-5182-6

Bontempi,G. et al (2012). Machine learning strategies for time series forecasting. *European Business Intelligence Summer School. Springer, Berlin, Heidelberg.*

Eco, R., Aquino, D., Lagmay, A., Alejandrino, I., Bonus, A., Escape, C., … Timbas, N. (2015). Landslide and debris flow susceptibility mapping of Leyte Province, Philippines using remote sensing, numerical modelling, and GIS. *Journal of the Philippine Geoscience and Remote Sensing Society*, *1*(1), 53–71.

Fawcett, T. (2005). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(2006), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association, 47(260), 583–621. https://doi.org/10.1080/01621459.1952.10483441

Lands Geological Survey Division. (2015). MGB's Geohazard Maps now viewable and downloadable online. Retrieved August 10, 2019, from http://mgb.gov.ph/en/2015-05-13-02-02-11/mgb-news/303-mgb-s-geohazard-maps-now-viewable-and-downloadable-online

McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, 51–56. Austin, Texas.

Nieva-Nishimori, A. (2018, September 18). "*Disaster imagination" to help prepare communities for disaster: Solidum*. Retrieved May 6, 2019, from ABS-CBN News website: https://news.abs-cbn.com/news/09/18/18/disaster-imagination-to-help-prepare-communities-for-disaster-solidum

Nikolopoulos, E. I., Borga, M., Creutin, J. D., & Marra, F. (2015). Estimation of debris flow triggering rainfall: Influence of rain gauge density and interpolation methods. *Geomorphology*, *243*, 40–50. https://doi.org/10.1016/j.geomorph.2015.04.028

Nolasco-Javier, D., Kumar, L., & Tengonciang, A.P.M. (2014). Rapid appraisal of rainfall threshold and selected landslides in Baguio, Philippines. *Nat Hazards*. https://doi.org/10.1007/s11069-015-1790-y

Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science Engineering*, *9*(3), 10–20. https://doi.org/10.1109/MCSE.2007.58

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Pham, B.T. et al. (2016). A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environmental Modelling & Software*, 85(2016), 240–250.https://doi.org/10.1016/j.envsoft.2016.07.005

scikit-learn developers (2019). *An introduction to machine learning with scikit-learn — scikit-learn 0.21.3 documentation*. Retrieved June 23, 2019, from https://scikit-learn.org/stable/tutorial/basic/tutorial.htm

Scheevel, C. R., Baum, R. L., Mirus, B. B., & Smith, J. B. (2017). *Precipitation thresholds for landslide occurrence near Seattle, Mukilteo, and Everett, Washington* (USGS Numbered Series No. 2017–1039; p. 60). Retrieved from U.S. Geological Survey website: http://pubs.er.usgs.gov/publication/ofr20171039

Welch, B. L. (1947). The generalization of 'student's' problem when several different population variances are involved. *Biometrika*, *34*(1–2), 28–35. https://doi.org/10.1093/biomet/34.1-2.28