Support Vector Regression on GIS and Map Data for Landslide Susceptibility of Laguna, Philippines

Ramon Leonardo C. Capunpon and Jaderick P. Pabico

***Abstract*— Given the province of Laguna's geographical situation, it is prone to different kinds of natural hazards. Landslide controlling factors are present in the area. This study attempted to create a landslide susceptibility mapping using an integrated weight index model by applying analytical hierarchy process (AHP) and frequency ratio (FR) on GIS and map based environmental data acquired from different government and non-government organizations. The features are extracted and merged using QGIS and Python scripts. The susceptibility model is then fitted into a Support Vector Regression model to find the correlation of the extracted factors to the respective landslide susceptibility ratings. A web-application was also develop to enable interaction with the resulting models.**

***Index Terms*— GIS,Python,AHP,FR,SVR,Machine Learning,landslideGIS,Python,AHP,FR,SVR,Machine Learning,landslideQ**

I. Introduction

Landslide is a movement of soil, rock, and debris down a sloped part of land. It is a natural disaster that sometimes is unavoidable when the stability of a slope decreases or changes due to natural or instigated factors. Basically, it is gravity acting upon the earth that already has pre-conditional factors. Landslide may be caused but not exclusive to these natural causes: erosion, earthquakes, piling of groundwater or flooding because of heavy rain, weak soil structure and composition. There are also human activities that may contribute to, or even cause, a landslide, these includes: deforestation, vibrations caused by heavy machinery, quarrying, and mining; with heavy rainfall, they will increase the fragility and decrease the integrity of the earth.

Landslide as a natural disaster can cause property damage, injury, and death. Multiple impact factors that interact and may cause landslides can be modeled to create a precusor signal for early warning and prevention of this natural disaster. Geographic Information Systems (GIS) along with maps and weather and precipitation data have been used with Machine Learning algorithms and predictive analytic models to make reliable and life-saving forecasts.

With the Philippine's aggressive modernization and infrastructure programs, and the abundance of elevated and sloped lands, paired with the increasing and intensifying typhoons, it is important to understand and map landslide susceptibility of its areas especially the provinces near elevated areas and bodies of water using a reliable prediction model.

II. Study Area

The province of Laguna in the Philippines, according to the Philippines Statistics Authority QuickStat for June of 2018, consists of 6 cities, 24 municipalities, and 681 barangays, with a total population of 3,035,081 and an annual growth rate of 2.47. The province is beside the southern part of Laguna de Bay, the largest lake in the Philippines and on its southern bother are two dormant volcanoes, Mount Makiling and Mount Banahaw. Cities and municipalities of the province are popular for numerous hot spring resorts found near the Mount Makiling slopes.

According to the Philippines Statistics Authority, the province of Laguna, Philippines, covers a total area of 1,917.85 km^2. It is situated at the southern part of Laguna de Bay. The province has 24 mountains, most of which are inactive volcanoes, including Mt. Banahaw giving the province the highest peak of 7,120 ft. It is also home to a significant number of tributaries due to its proximity to Laguna de Bay, Taal Lake, Caliraya Lake, and Lumot Mahipnon Lake.

Some cities and municipalities of the province of Laguna are along the Marikina Valley Fault system; being situated in areas where landslides factors are present and can inflict great disaster, the lack of studies and data on landslide susceptibility in the province of Laguna poses a great threat to its residents.

III. Review of Related Literature

In a study by Neenu and Lakshm [1]i in 2016, they created a prediction model for landslide in Cherrapunjee region of Meghalaya using Support Vector Machine(SVM), a technique meant to have strong capability to predict landslide by forecasting rainfall dataset on rainfall analysis. They concluded that SVM proved to be an efficient technique to forecast landslide by predicting the rainfall in advance.

Poonam and Neelam [2] as cited in a Journal in 2018, enhanced the study of landslide prone area in Varunavat Parvat, India through supervised analysis. The area is located beside the Himalaya, situated at the right side of river Bhagirathi at a height of 1150m. They prepared data including slope instability and amount of rainfall to create prediction models using Artificial Neural Networks (ANN), Support Vector Machines (SVM), and logistic regression. They classified the outcomes of their predictions into classes being low, medium, or yes, no. Their comparison shows that SVM perform well in terms of accuracy for considered factors and in some cases logistic regression is performing better than SVM but SVM model best fits the hyperplane which divide the landslide prone groups.

In 2015, Poonam, Vibhuti, and Shivani [3] created a GIS based model for monitoring and prediction of landslide susceptibility. They pointed out in their study that landslides tend to occur at any point in time and can cause huge damages to human life and resources, but the advancement in GIS based applications has eased out working on spatial or geographical data and will provide a powerful tool to model the landslide hazards for their spatial analysis and prediction. They proposed a GIS based landslide monitoring and forecast system using sensors. Using k-means clustering and ID3 decision trees they proposed a system to efficient and timely generation of landslide alerts.

In 2009, Milos and Branislav [4] used SVM and k-Nearest

Presented to the Faculty of the Institute of Computer Science, University of the Philippines Los Baños in partial fulfillment of the requirements for the Degree of Bachelor of Science in Computer Science

Neighbor(k-NN) algorithms trained upon expert based model of landslide susceptibility using a multi-criteria analysis. They weighted the influences of different input parameters using Analytical Hierarchy Process (AHP). Their parameters included elevation, slope angle, aspect, distance from flows, vegetation cover, lithogy, and rainfall to represent the natural factors of the slope stability. The study, using machine learning classifiers included pattern recognition algorithms performed through training and testing mode. The SVM classifier outperformed the accuracy of the k-NN and turned out as quite a convenient classifier for landslide susceptibility as it turned out to be more consistent and precise. The most important result in their study was revealing that small training sets are sufficient to reach very high accuracy. They proposed to use a wider area but a multi-fold case with sparse inputs is yet to be confirmed.

In 2018, Jiubin, Yuanxue, and Ming [5] performed a study on optimisation algorithm for decision trees and the prediction horizon displacement of landslide monitoring. In their study, they have pointed out the feature importance of the different attributes from GIS. They created a feature selection model which pointed out the important features: monitoring point, water level drop speed, monthly rainfall, max water level, daily rainfall, mean temperature, extreme minimum pressure, extreme minimum temperature, etc.

In 2017, Shuangxi, Qing et. al. [6] created a model for predicting the landslide deformation with a knowledge-guided approach based on multi-mode monitor data using a Support Vector Regression (SVR). Using sensitivity coefficients to reflect the sensitive degree induced by multiple influencing factors, then a k-means clustering was implemented to discover the mechanism knowledge rules and finally deformation was predicted using SVR under the guiding of priori rules. They have concluded that their proposed knowledge-guided SVR approach is superior to the conventional SVR as verified in comparative experiments using displacement monitoring GIS datasets.

IV. Data Acquisition

A. Data Layers

In order to map the landslide susceptibility of the province of Laguna, data from different local and international; government and non-government organizations were acquired. (1)Slope data and (2)land cover were acquired from the Philippines' National Mapping and Resource Information Authority which came in the form of shape files, and acquired through a formal data request. (3) Elevation data in the form of geotiff file - ASTER GDEM was acquired from PhilGIS and NASA's Earth Observing System(EOS) through the Landviewer web application that automatically accesses United States Geological Survey (USGS) and Shuttle Radar Topography Mission (SRTM) databases. (5) Near-Infrared and Red Bands were also acquired from EOS Landviewer. Vector maps of Philippines' (6)water ways, (7)water basins were acquired from Geofabrik, a company that provides free vector shape data derived from OpenStreetMap. (8) Lithology data was acquired from the Bureau of Soils and Water Management(BWSM), PhilSoil project, and downloadable data in PhilGIS, (9) Landslide related hazard ratings(liquifaction, earthquake-induced, rain-induced) were acquired from Philippine Institute of Volcanology and Seismology(PHILVOCS) , and (10) rainfall and weather data was acquired from the Philippine Atmospheric, Geophysical and Astronomical Services Administration(PAGASA).

1) Landslide Inventory: **The landslide inventory is essential in the assessment of landslide susceptibility. It is the basis of the study and will be used in the creation of the susceptibility model, because the factors present in the landslide areas are expected to be present on susceptible and critical areas. To derive a landslide inventory, two techniques were used: (1) NASA Drip-Slip algorithm using Earth Engine. Drip-Slip is a landslide identification and extreme precipitation monitoring software that detect spectral changes from Landsat imagery that may indicate landslides. (2) Practical identification of Landslides using Google Earth Pro as proposed by M. Mihir and B. Malamud where areas are scanned for coloration and characteristics of landslide or land movement.**
This study focused on the landslide areas prevalent during the year 2015 to 2016 because most of the data acquired from the government offices are of the said date range. With the help of QGIS, a total of 319 suspected landslide areas were determined and mapped.

TABLE I
Data layers from the study area that were used in the study

Data Layer	**Format**	**Data Source**
GDEM	Geotiff, Raster	ASTER, PhilGIS, EOS Landviewer
Landsat 7	Geotiff, Raster	PhilGIS, EOS Landviewer
Landsat 8 (NIR, NR)	Geotiff, Raster	PhilGIS, EOS Landviewer
Slope	Shapefile	NAMRIA
Land Cover	Shapefile	NAMRIA, PhilGIS
Land Use	Shapefile	NAMRIA, PhilGIS,Geofabrik
Lithology	Shapefile	NAMRIA, PhilGIS (PhilSoil, BSWM)
Water ways and Basins	Shapefile	Geofabrik
Landslide maps	Geotiff	Earth Engine, Google Earth Pro
Rainfall	CSV	PAGASA

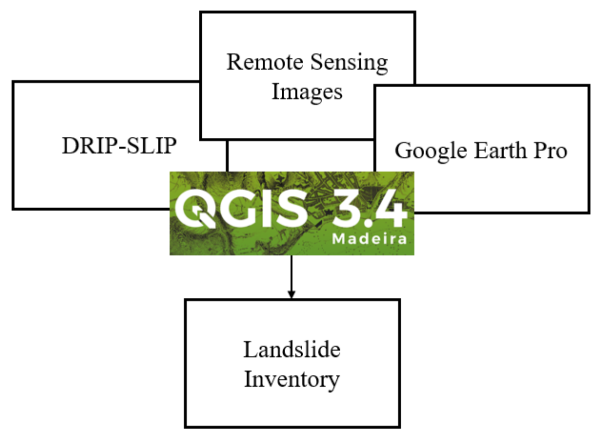

Fig. 1. Source of Landslide Inventory

2) Landslide Controlling Factors: **Landslides are results of different triggering factors, including, but not limited to, geological, anthropogenic and meteorological factor that can trigger land movement down a slope; these are referred to as landslide controling factors [7]. The Mines and Geoscience Bureau, the organization that provides landslide susceptibility data for the country uses a limited number of factors in the creation of their susceptibility map, namely: (1)Lithology (Rock Type), (2)Degree of Weathering (Ground Integrity), (3)Degree of Fracturing and (4)Orientation of Structures (Aspect). Their in-house geologists, perform field work to gather these valuable data to generate the maps. There are no existing standard or guideline in the creation of landslide hazard maps, but the scale of analysis, the nature of the study area, the data availability, and the quasi-empirical and statistical criterions in literature can be referenced [8]. A question was raised during the consultation**

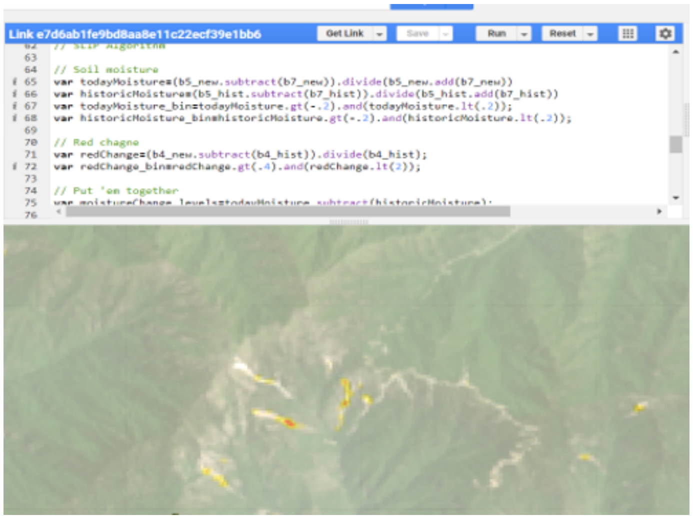

Fig. 2. Screen shot from Earth Engine where the Drip-Slip algorithm was implemented

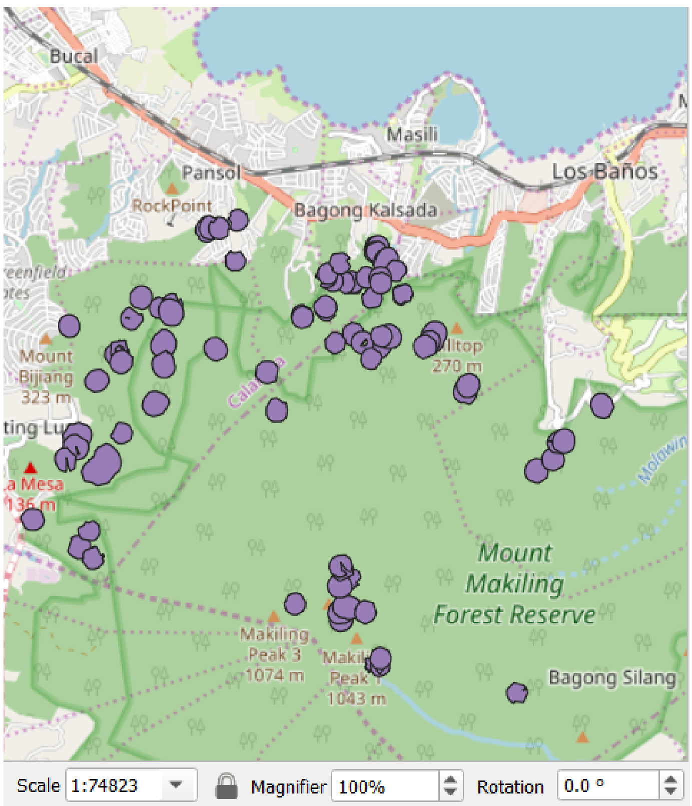

Fig. 3. Potential landslide areas vectorized and buffered in QGIS

with government agencies that landslide susceptibility is greatly affected by the amount of rainfall an area is receiving. Thus, this study included rainfall in the considered landslide controlling factors.

In this study, slope, aspect, elevation, ground integrity, distance from faults, distance from waterways, land use–land cover (LULC), normalized difference vegetation index (NDVI), and Terrain Ruggedness Index(TRI) were selected as the static landslide controlling factors and rainfall as a dynamic landslide controlling factor.

Slope is one of the most recognized topographic landslide controlling factor which significantly affect the probability of a landslide(Ayalew and Yamagishi, 2005; Chalkias et al., 2016). When the slope reaches a gradient higher than 15 degrees, the possibility of a landslide happening increases (Lee and Min, 2001). The slope data of the study area was provided by NAMRIA in the form of shape files, these shape files used polygons to map the slope. Using QGIS, centroids were extracted to get the potential points of interests, totaled to almost half a million data points (445,000) . There data points contain slope data and coordinates that will be used to extract the other landslide controlling factors from all the collected data formats.

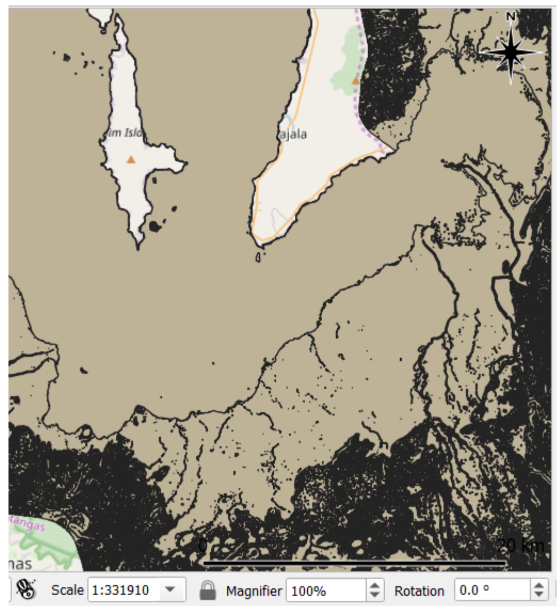

Fig. 4. Slope layer acquired from NAMRIA processed in QGIS

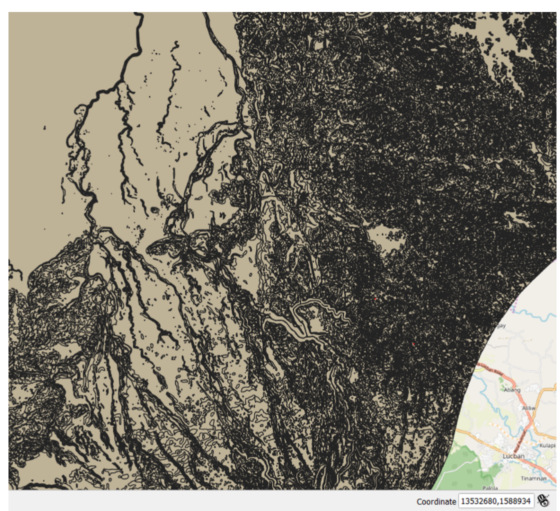

Fig. 5. Slope layer acquired from NAMRIA processed in QGIS in larger scale

Aspect refers to the direction where the slope faces, is related to the amount of water a soil type can hold, ruggedness of terrain and vegetation, which indirectly affects landslide development [9]. Aspect was derived from the elevation map raster using QGIS geometry processing functions. The aspects were initially

in accurate degrees, the data was transformed into directional classifications.

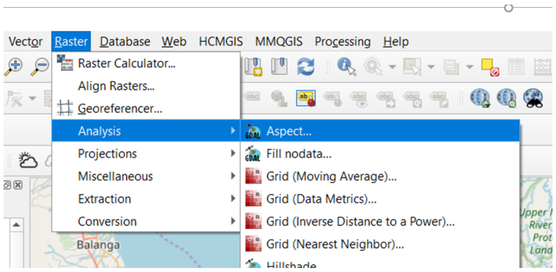

Fig. 6. Calculating Aspect using QGIS Raster functions

Elevation, is the measure of land surface height, it is one of the key factors in determining gravitational potential energy of terrain [10]. Multiple studies have considered elevation as one of the most important landslide controlling factors. In the study area, the highest elevation is located within the Mt. Banahaw area, reaching a peak of 2,170 m (7,120 ft). Elevation can be extracted from raster files readily downloadable from multiple open source tools that archives satellite images. This study used Digital Elevation Maps from the Landsat 7 and 8 programs that were collected from the PhilGIS and EOS Landviewer websites, the most recent elevation data were used in this study.

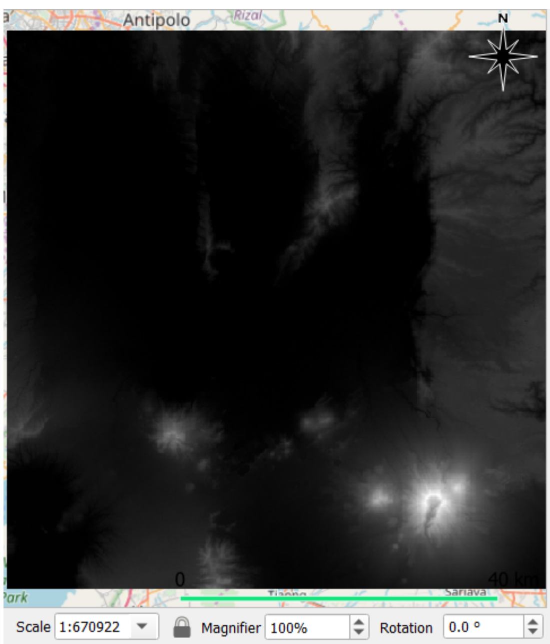

Fig. 7. Elevation Layer Landsat 7 ASTER DEM imported in QGIS

Lithology is directly related to the slope stability, as the slope gets steeper, the lithological attributes will greatly determine if a landslide will be triggered (Guo et al., 2015). The ability of the soil to hold water and prevent breaking also hold key ingredients in the possibility of landslides. The cohesion factor will determine if the integrity of the soil is enough to hold significant amount of stress at a specific slope. Ground Integrity was derived from the lithology and soil data acquired from the PhilSoil project data provided through the PhilGIS downloads, lithology from NAMRIA, and soily data from BSWM. Converging these different data sets were a challenge, but was crucial in improving the richness of the data set. Land Use and Land Cover is also a direct factor affecting the possibility of landslides [11].

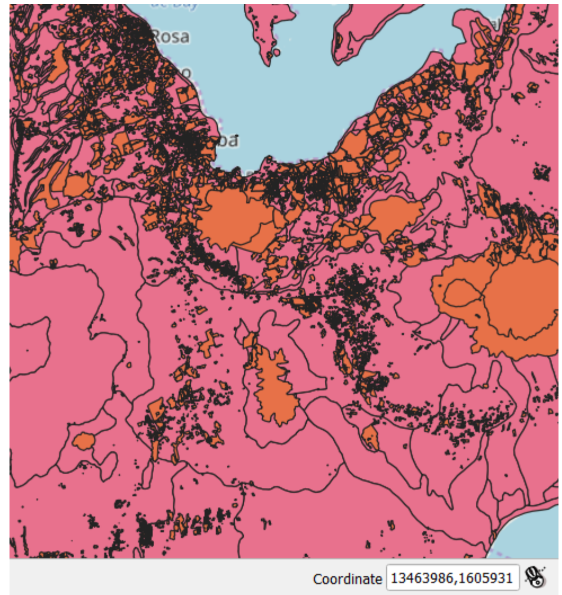

Fig. 8. Mergin of Lithology layers from PhilSoil, PhilGIS, and Geofabrik using QGIS

The susceptibility of landslides increases during tectonic movement, thus distance to faults is also a landslide controlling factor. Earthquake triggered landslides are normally found in areas near a fault, distances of a slope from geological tectonic zones are important to consider in slope analysis (Fan et al, 2018). Distance to water ways and basins are also a factor to consider in slope analysis and landslide prediction [12]. Water moving along a steep slope contribute to erosion and terrain movement. In this study, the faults were mapped using shapefiles provided by PHILVOLCS and open-source data provided by Geofabrik.

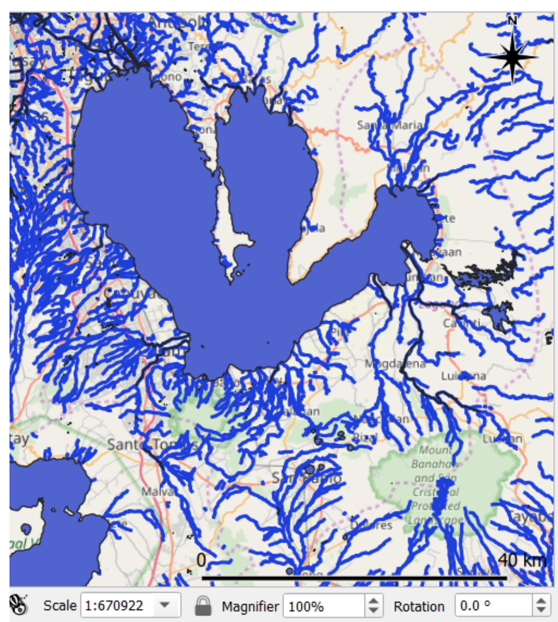

Fig. 9. Water ways and water basin layers from Geofabrik merged together using QGIS

Landslide can be indirectly affected by vegetation through the prevention of soil erosion. NDVI is the measurement of vegetation coverage using the observation of the visible red and near-infrared regions. The NDVI is calculated using the following

equation:

$$\frac{\mathbf{NDVI} = \mathbf{DN}_{NIR} - \mathbf{DN}_{R}}{\mathbf{DN}_{NIR} + \mathbf{DN}_{R}} \tag{1}$$

Where DIN$_{NIR}$ stands for the spectral refletance derived from the measured radiances in the near-infrared regions(NIR), and DN$_{R}$ stands for the sprectral reflectance derived from the measured radiances in the visible red regions. In this study, the red and near-infrared basebands where downloaded from EOS Landviewer and then processed in QGIS using a python script that executes the above equation.

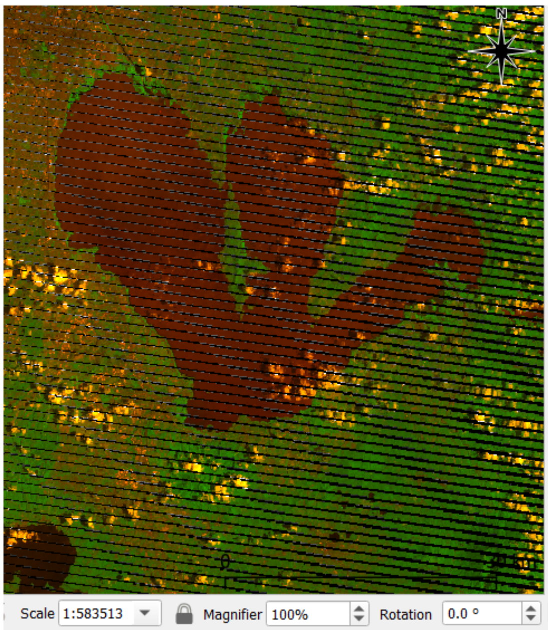

Fig. 10. Combined Visible Red and Near-Infrared bands from Landsat 7 data layers

Another factor that supports the relationship of slope and elevation is the terrain ruggedness.”Terrain Ruggedness Index(TRI) is a measurement developed by Riley, et al. (1999) to express the amount of elevation difference between adjacent cells of a digital elevation grid. The process essentially calculates the difference in elevation values from a center cell and the eight cells immediately surrounding it. Then it squares each of the eight elevation difference values to make them all positive and averages the squares. The terrain ruggedness index is then derived by taking the square root of this average, and corresponds to average elevation change between any point on a grid and its surrounding area.” - https://www.edenextdata.com/?q=content/terrain-ruggedness-index-tri

V. Methodology

In this study, an landslide susceptibility model was built by using Analytical Hierarchy Process and Frequency Ratio, and then used to create a Support Vector Regression model based on the features showing the heaviest feature weights. The model is then integrated in an interactive web application to provide functionality to incorporate rainfall data. The web application also provides a way to access the regression model.

1) Feature Cleaning and Transformation: The collected data was not ready for modeling initially. Features had to be mapped to numerical values and cleaned; removal of empty values and extreme outliers were done. Some features were also normalized to create a better representation of the data. From the starting feature count of 440,000 points derived from the slope data layer

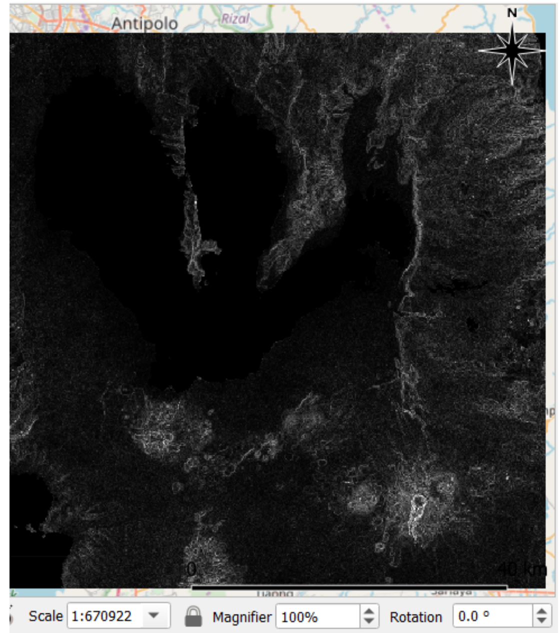

Fig. 11. Computed Terrain Ruggedness Index from Landsat DEM raster using QGIS

by extracting the centroids of each slope polygon layer, the final feature count was 235,786 data points. The data points were then merged with the landslide inventory to create the final data set.

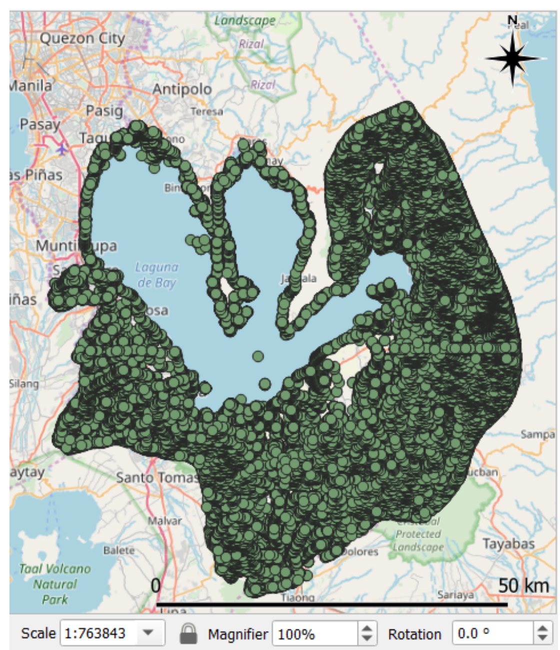

Fig. 12. Centroid points derived from the slope layer that will be used to merge all data layers and feature extraction

The ground integrity is the product of cohesion factor(C_i) and water capacity(WC_i) of the ground. With the formula:

$$GI_i = C_i \times WC_i \tag{2}$$

A. Analytical Hierarchy Process(AHP)

The AHP is a multi-criteria decision-making method developed in the late 1970’s by Thomas L. Saaty (Saaty, 1977). It is a method

TABLE II
ASPECT CODE AND THEIR EQUIVALENT VALUE IN DEGREES

Aspect Direction Code	**Direction (degrees)**
N	0
NNE	0.0225
NE	0.045
ENE	0.0675
E	0.09
ESE	0.1125
SE	0.135
SSE	0.1575
S	0.18
SSW	0.2025
SW	0.225
WSW	0.2475
W	0.27
WNW	0.2925
NW	0.315
NNW	0.3375

TABLE III
SLOPE TRANSFORMATION TO NUMERICAL VALUE FOR EASIER CLASSIFICATION

Slope Type	**Ceiling Value**
Flat(0-8%)	.8
Gentle(8-18%)	.18
Downhill(18-30%)	.3
Steep(30-50%)	.5
Very Steep(greater than 50%)	1

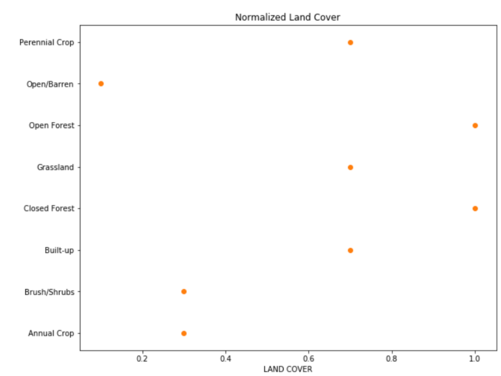

Fig. 13. Normalized land cover values

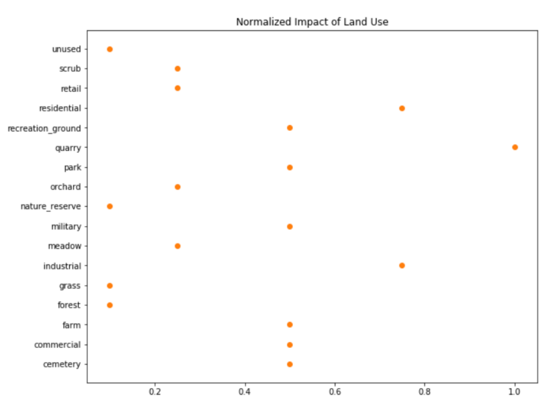

Fig. 14. Normalized land use impact values

to derive ratio scales from paired comparisons. Experts' opinions

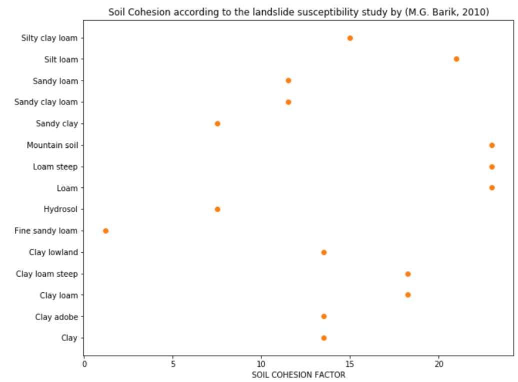

Fig. 15. Soil cohesion factor for soil classes present in the data according to a study by M.Barik, 2010

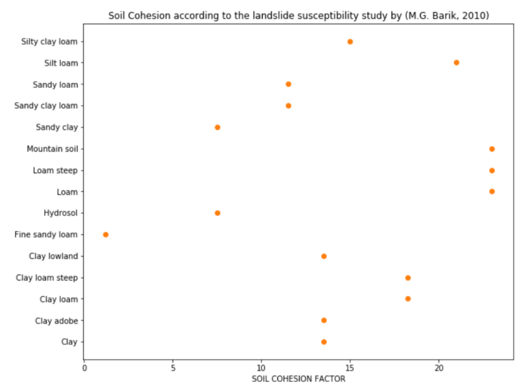

Fig. 16. Soil water capacity of soil classes present in the data according to the United States Geological Survey

and knowledge are used in order to get the relative importance of each factor.

A nine-point scale is used to compare the importance of the complex factors. A score of 1 to 9 is given when a factor is more important than the other and when it's the opposite, the score given is between 1/2 and 1/9. The higher the score, the greater the importance of the factor. In order to check the consistency of the pair-wise comparison matrix, the computed consistency ratio must be less than 0.1.

The pair-wise comparison matrix used in this study is based on a study by Yi, Zhang et. al., in 2019 they created pair-wise matrix for the landslide susceptibility mapping for Jiuzhaigou region of Sichuan Province, China. The matrix was then subjected to consultation from geologist from the Mines and Geosciences Bureau of the Philippines. The consistency ratio of this study, given 10 factors is 0.02967 which satisfies that it must be below the 0.1 consistency requirement.

The pair-wise matric created in the study is shown in Fig. 17.

B. Frequency Ratio

The Frequency Ratio method is used to determine the relationship between the landslide locations and the landslide controlling factors. The assumption is that, a landslide will most likely occur with same conditions as the previously occurred landslides. It is the ratio of probability of occurrence to non-occurrence of the different landslide factors. The FR value is calculate using the formula:

$$FR_i = \frac{Npoint(S_i)/Npoint(N_i)}{\sum Npoint(S_i)/\sum Npoint(N_i)} \tag{3}$$

Factor	a(1)	a(2)	a(3)	a(4)	a(5)	a(6)	a(7)	a(8)	a(9)	a(10)	Weight
Elevation(a1)	1.0	0.25	2.0	0.33	0.25	1.0	0.33	0.5	2.0	0.33	0.051470
Slope(a2)		1.00	4.0	2.00	1.00	3.0	2.00	3.0	4.0	1.00	0.182471
Aspect(a3)			1.0	0.33	0.25	0.5	0.33	0.5	1.0	0.25	0.036257
Ground Int. (a4)				1.00	0.50	1.0	0.50	2.0	3.0	0.50	0.097347
Fault Distance(a5)					1.00	2.0	1.00	3.0	4.0	0.50	0.155214
LULC(a6)						1.0	0.50	1.0	2.0	1.00	0.078171
TRI(a7)							1.00	2.0	3.0	0.50	0.126687
Water Distance(a8)								1.0	2.0	1.00	0.076012
NDVI(a9)									1.0	0.33	0.037505
Rainfall(a10)										1.00	0.158866
Consistency Ratio: 0.0296770											

Fig. 17. The pair-wise comparison matrix, factor weights, and consistency ratio obtained in the study

where $Npoint(S_i)$ represents number of points recognized as landslides in class i, and $Npoint(N_i)$ represents total number of points belonging to class i in the whole area. $\sum Npoint(S_i)$ stands for the total number of points recognized as landslides in the whole area, and $\sum Npoint(N_i)$ represents total number of points in the whole area.

Factor	Class	FR	Weight	Factor	Class	FR	Weight
Elevation	0.05	2.1865815	0.051470	Slope	0.5	1.577908	0.182471
	0	1.6391313			0.3	1.4246328	
	0.1	0.8740408			0.08	0.9702526	
	0.15	0.4670532			0.18	1.0715143	
	0.3	0.5846829			1	2.1960833	
	0.25	1.6672571			0	0	
	0.2	1.1510644		Ground Int.	0.354	0.9643743	0.097347
	others	0			0.474	8.7150209	
Fault Distance	0.085	1.3739074	0.155214		0.5688	0.5073335	
	0.17	1.5075443			1	0.8961351	
	0	1.2534067			0.72	0.5255032	
	0.2551	0.1582373			0.24	13.635126	
					0.28	1.3802388	
					others	0	

Fig. 18. Normalized frequency ratio and weight for each class of each feature in the data set (Elevation, Slope, Ground Integrity, and Fault Distance)

Factor	Class	FR	Weight	Factor	Class	FR	Weight
Aspect	0.1575	0.935	0.036257	Water Dist.	0.0152	0.7551	0.076012
	0.0675	1.4303			0	1.3368	
	0.27	0.6653			0.0455	1.7151	
	0.09	1.9839			0.0303	0	
	0.2475	0.7588					
	0.135	1.5298					
	0.225	0.775		NDVI	0.4	1.5828	0.037505
	0.315	1.432			1	1.0226	
	0.1125	1.2252			0	0.7866	
	0.2025	1.4654			0.1	0.8807	
	0.045	1.4384		Rain	0.2	0.1	0.158866
	0	1.4964			0.4	0.25	
	0.2925	0.9447			0.6	0.5	
	0.3375	1.3035			0.8	0.75	
	0.18	1.0749			1	0.1	
	0.0225	1.4099					

Fig. 19. Normalized frequency ratio and weight for each class of each feature in the data set (Aspect, Water Distance, NDVI, Rainfall)

Factor	Class	FR	Weight	Factor	Class	FR	Weight
LULC	0.3	1.2931	0.078171	TRI	0.55	1.0845	0.126687
	1	3.1865			0.5	1.4127	
	0.7	0.9907			0.6	1.1816	
	0.06	0.9223			0.45	1.3472	
	0.02	65.1727			0.7	1.1469	
	0.04	0.1075			0.4	1.468	
					0.65	1.1726	
					0.35	1.9397	

Fig. 20. Normalized frequency ratio and weight for each class of each feature in the data set (LULC and Terrain Ruggedness Index)

TABLE IV

CLASSIFICATION OF LANDSLIDE SUSCEPTIBILITY USING MEAN AND STANDARD DEVIATION ON THE INTEGRATED WEIGHTED INDEX.

Class	**Range**
None - Very Low	less than 0.64
Low	0.64 - 0.87
Moderate	0.88 - 0 1.11
Relatively High	1.12 - 1.135
High	1.36 - 1.59
Very High	1.6 - 1.8
Hazardous	greater than 1.8

C. Integrated Weight Index

The AHP method is used to identify the mutual relationship between the landslide controlling factors, but is heavily subjected to the experts' knowledge and literature. Meanwhile, the FR method can show the influence of each controlling factor by counting their contributions. By combining the two methods, both the mutual relationship and the contributions are considered. The integrated weighted index measures the probability of a landslide by combining the FR and AHP methods. The integrated weighted index is calculated using the formula:

$$I = \sum_{m}^{i} (W_i \times FR_i) \tag{4}$$

where *m* stands for the number of controlling factors, N_i is the weight of each controlling factor calculated by the AHP method and $FR_i$is the calculated FR value of the controlling factor. Acquiring the mean and the standard deviation provides a way for classification of landslide susceptibility.

In this study, the integrated weighted index formula was used to calculate the susceptibility of each potential landslide point. By acquiring the mean, which is 1.12, and the standard deviation, 0.24 of the integrated weighted index of the data set, landslide susceptibility was classified into seven classes:

D. SVR

In Support Vector Regression is a generalization of the Support Vector Classification, in which the model returns a continuous-valued output, as opposed to an output from a finite set. In other words, a regression model estimates a continuous-valued multivariate function (Awad and Rahul, 2015)

One of the goals of this study is to create a regression model where a user can easily calculate the susceptibility rating of a presented area with only a portion of the features used in creating the susceptibility model derived from the integrated weighted index.

The study wanted to know if the susceptibility can still be derived given only half of the landslide controlling factors. The

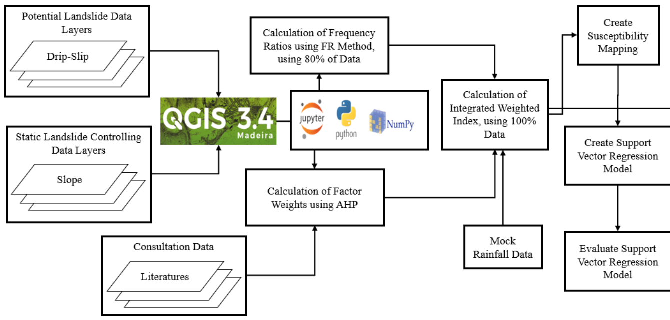

Fig. 21. Flowchart of the creation of the landslide susceptibility model and application of the Support Vector Regression Model)

TABLE V
WEIGHTS OF EACH LANDSLIDE CONTROLLING FACTOR BASED FROM THE AHP METHOD IN DESCENDING ORDER

Feature	Weights
Slope	0.182471
Rainfall	0.158866
Fault Distance	0.155214
TRI	0.126687
Ground Integrity	0.097347
LULC	0.078171
Water Distance	0.076012
Elevation	0.05147
NDVI	0.037505
Aspect	0.036257

features with the highest weights were used in training the machine learning model namely: Slope, Rainfall, Fault Distance, TRI, and Ground Integrity. The sorted list of the weights of the landslide controlling factors presented on Table V.

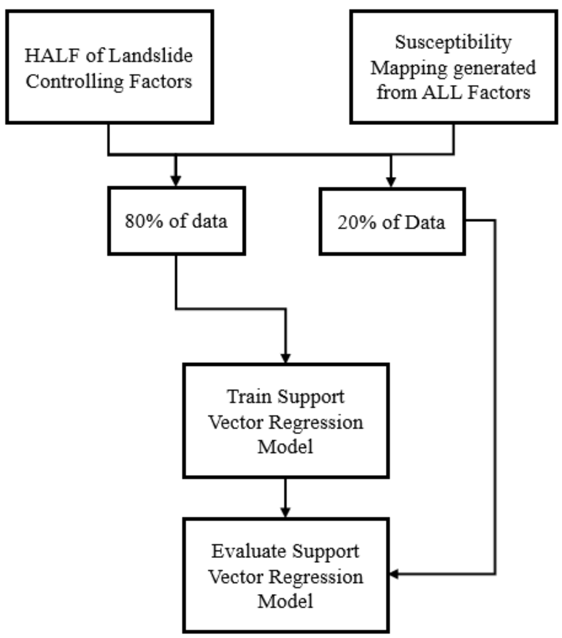

Fig. 22. Flowchart of the creation and evaluation of the Support Vector Regression Model

Using the Support Vector Regression library from Scikit Learn, a free software machine learning library for the Python programming language. The regression model was created and continuously improved by using StratifiedShuffleSplit and GridSearchCV on the model parameters from the model selection library.

E. Web Application for Dynamic Controlling Factor

In this study, Rainfall was considered as a dynamic landslide controlling factor, a factor that can change indefinitely. In order to visualize the landslide susceptibility efficiently, a web application was developed.

The web application serves as the tool to dynamically supply the rainfall data to the existing model created using the static controlling factors. A grid view is overlayed on a map of the study area, each grid can be toggled to increase or decrease the amount of rainfall. The rainfall data of each point is updated and used to recalculate the susceptibility.

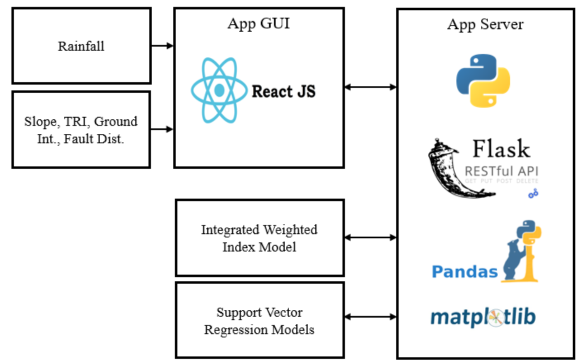

Fig. 23. High-Level Architecture of the Web Application

The web application was also created for using the support vector regression models that were created by this study.

VI. RESULTS AND DISCUSSIONS

The main purpose of this study is to utilize Support Vector Regression in a landslide susceptibility model generated using AHP and Frequency Ratio and to be able to apply rainfall data as a dynamic landslide controlling factor to get a dynamic landslide susceptibility mapping.

The usage of AHP and FR methodologies created a complementing approach to the creation of an integrated weighed index for landslide susceptibility, and in turn paved a way to the successful application of the Support Vector Regression algorithm.

By analyzing the results of applying the integrated weighted index model for the controlling factors and rainfall on all the known potential landslide areas, we can see that a very small fraction of all the landslide points were given a low rating. More than half the points rated as moderate have integrate weighted values of are greater than 0.95.

The result was expected given the variety of data in the landslide inventory. Nevertheless, the landslide inventory is generally incomplete, and is affected by many factors, such as the quality and scale of remote sensing images, the effectivity of the tools used, and the expertise of the interpreter involved (Malamud et al., 2004). Improving the landslide inventory will greatly improve the performance of the model.

A. Performance of the Support Vector Regression Model

In order to train the regression model, the whole data set was split into two: 80% training and 20% testing using the train_test_split and StratifiedShuffleSplit module from

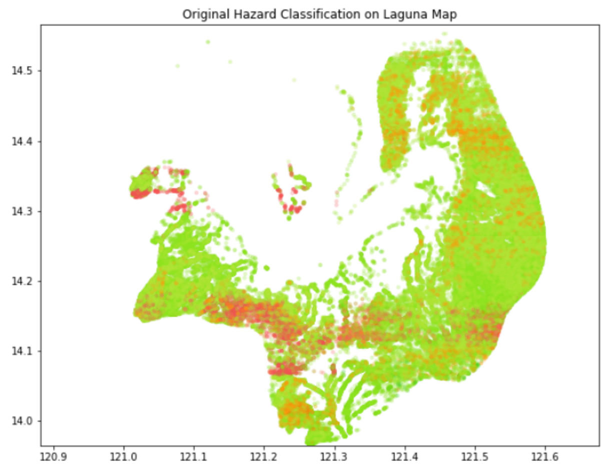

Fig. 24. Integrated weighted integration index using the static landslide controlling factors

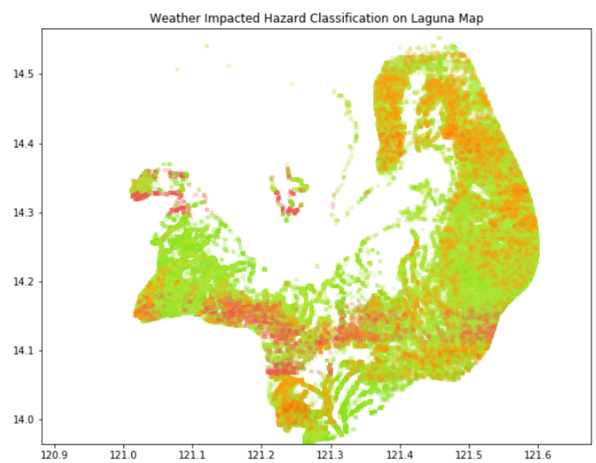

Fig. 25. Adjusted weighted integration index using rainfall data

sklearn.model_selection library. The regression model was trained iteratively to get the highest possible performance. Performance of the support vector regression model was measured using rsquared and mean squared error.

R-squared is a statistical tool that shows the proportion of the variance for a dependent variable explained by an independent variable or variables in a regression model. It is generally the percentage of landslide susceptibility that can be explained using the given landslide controlling factors.

$$r^2 = 1 - \frac{Explained\,Variation}{Total\,Variation} \tag{5}$$

The mean squared error shows how close the regression line is to the data points by calculating the distances from the points to the regression line and squaring them. The squaring results to no negative values and more weights to larger values. It is the average of the set of all the errors.

The machine used in the creation of the model posed a challenge on the resources needed to create an acceptable regression model. Given the varying results and the drastic differences on results when using different kernels, a result of 88.9 percent was a breakthrough. This means that more than 88 percent of the landslide susceptibility results can be predicted by only providing the top 5 landslide controlling factors.

TABLE VI

DISTRIBUTION SUSCEPTIBILITY OF KNOWN POTENTIAL LANDSLIDE AREAS

Susceptibility Rating	Count	Adjusted Count (Heavy Rain)	Percentage
None to Very Low	0	0	0
Low	37	25	0.008
Moderate	1275	1232	0.41
Relatively High	244	295	0.098
High	8	12	0.004
Very High	212	160	0.053
Hazardous	1217	1269	0.424
Total	**2993**		

TABLE VII

THE BEST R-SQUARED AND MEAN SQUARED ERROR(MSE) VALUES ACQUIRED ON THE TEST DATA SET USING DIFFERENT KERNELS

Kernel	C	gamma	r-squared	mean squared error
Linear	3	1	0.1172	0.8942
RBF	5	2	0.8896	0.1091
Poly	100	auto	0.1849	0.8131

B. Web Application

Using React.js for the front end and Python with Flask on the back end, the web application was successfully integrated to the integrated weighted index and support vector regression model.

Functionalities provided by the web application: 1. Displays the current landslide susceptibility of the province of Laguna over a satellite map

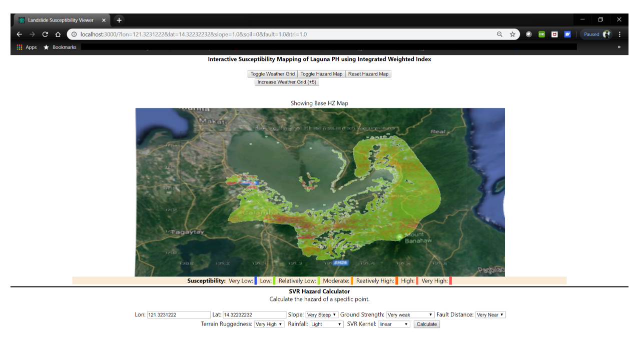

Fig. 26. View of the web application Graphical User Interface showing the static landslide susceptibility of the province of Laguna cakculated using the static landslide factors

2. Provides an interactive weather/rainfall interface to change the rainfall value of the susceptibility model.

3. Provides an form to input the 5 landslide controlling factors: Slope, Ground Integrity, Terrain Ruggedness Index, Distance from Faults, and Rainfall that will be used by the Support Vector Regression Model to predict the landslide susceptibility

REFERENCES

[1] N. Rachel and M. Lakshmi, *Landslide prediction with rainfall analysis using support vector machine*, June 2016, vol. 9.

[2] P. Kainthura and N. Sharma, "Enhancing the study of landslide prone area through supervised analysis: a case study of varunavat parvat, uttarkashi, india." [Online]. Available: 2018

[3] P. Kainthura, V. Singh, and S. Gupta, "Gis based model for monitoring and prediction of landslide susceptibility," 2015.

[4] M.Marjanovic, B.Bajat, and M.Kovacevic, "Landslide susceptibility assessment with machine learning algorithms," 2009.

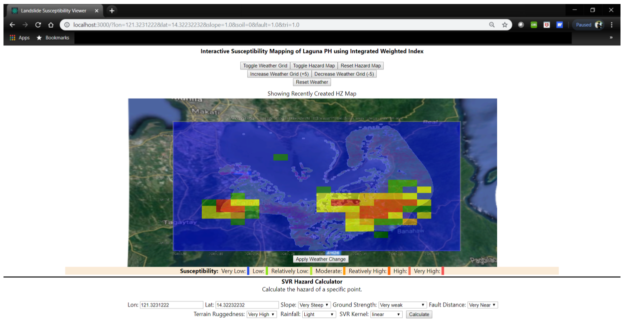

Fig. 27. View of the web application Graphical User Interface showing the rainfall adjustment grid

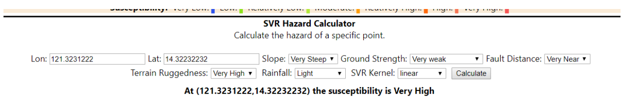

Fig. 28. View of the web application Graphical User Interface showing the form to get the landslide susceptibility rating of a specific point using the Support Vector Regression Model

[5] J. Zhao, Y. Liu, and M. Hu, "Optimisation algorithm for decision trees and the prediction of horizon displacement of landslides monitoring," 2018.

[6] S. Miao, Q. Zhu, L. Zhang, C. L. B., Zhang, and M. Chen, "A knowledge-guided landslide deformation prediction approach based on svr," 2015.

[7] B. S.-B. Wang, J. Lü, G.-N., Zhou, P.-G. Hou, S.-S., and S.-N. Xu, "Gis-based logistic regression for landslide susceptibility mapping of the zhongxian segment in the three gorges area, china, geomorphology," pp. 115, 23–31, 2010. [Online]. Available: https://doi.org/10.1016/j.geomorph.2009.09.025

[8] C. Romer and M. Ferentinou, "Shallow landslide susceptibility assessment in a semiarid environment – a quaternary catchment of kwazulu-natal, south africa," pp. 201, 29–44, 2016. [Online]. Available: https://doi.org/10.1016/j.enggeo.2015.12.013

[9] B. Q., C. Y., D. S., W. Q., Y. J., and J. Zhang, "An improved information value model based on gray clustering for landslide susceptibility mapping," pp. 6, 18, 2017. [Online]. Available: https://doi.org/10.3390/ijgi6010018

[10] C. M., S. Pascale, R. G., and S. F., "Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the turbolo river catchment (northern calabria, italy)," pp. 113, 236–250, 2014. [Online]. Available: https://doi.org/10.1016/j.catena.2013.08.006

[11] M. R. Mansouri Daneshvar, *Landslide susceptibility zonation using analytical hierarchy process and GIS for the Bojnurd region, northeast of Iran, Landslides*, 2014.

[12] K. D. P., A. M. K., S. S., and G. R. P, "A comparative study of conventional, ann black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in darjeeling himalayas," 2006. [Online]. Available: https://doi.org/10.1016/j.enggeo.2006.03.004