1. USGS Award Number: G22AP00048

2. Project Title: **Innovative Data-Driven Frameworks for Geospatial Ground Failure Models**

3. Principal Investigator:

Laurie G. Baise, Ph.D.
Tufts University
200 College Ave
Medford, MA 02155
Tel: (617) 627-2211, Fax: (617) 627-3994
Laurie.baise@tufts.edu

Babak Moaveni, Ph.D.
Tufts University
200 College Ave
Medford, MA 02155
Tel: (617) 627-3211, Fax: (617) 627-3994

Term covered: January 1, 2022 – December 31, 2022

"This material is based upon work supported by the U.S. Geological Survey under Grant No. G22AP00048."

**Abstract:**
In recent years, the USGS and others (including the PI of this effort) have worked to develop predictive regional models for ground failure with a focus on landslides and liquefaction. The models provide probability estimates of ground failure given the shaking from an earthquake event. The current USGS implementation includes the ground failure models as a product on the overview page for each earthquake on the USGS Earthquake Hazard Program website. Many of the existing geospatial ground failure models have been developed using logistic regression. In this project, we will demonstrate three innovations in model development: 1) Uncertainty quantification for geospatial liquefaction logistic regression models; 2) Machine learning algorithms for geospatial liquefaction models; and 3) Machine Learning algorithms to create landside inventories using post-event imagery. The geospatial liquefaction model as developed by Zhu et al (2015; 2017) and since expanded by Rashidian and Baise (2020) and Baise et al. (2021) relies on geospatial proxies for soil density and soil saturation combined with earthquake loading estimates from USGS ShakeMap to predict the spatial extent of liquefaction after an earthquake.

**Significant contributions of the proposed research to USGS Earthquake Hazards Program**
The proposed research most directly addresses the following **program elements** of the External Research Program Announcement for FY 2022: **Element I**: as it updates "National and regional assessments of the expected degree of ground deformation.", and **Element III:** specifically, "with the goal of improving hazard assessments". This research is also relevant to the following common priority topic for all research areas: "Develop approaches to provide earthquake hazard information needed for risk assessments." The proposed work is inherently aligned with the following Priority Topic for Research on **Engineering Seismology and Impacts (ESI):** "Improve predictive models of earthquake-triggered ground failures including landslides and liquefaction that can be characterized and employed at regional/global scales rather than site-specific or material behavior studies." And "The ESI panel prioritizes innovative models." And "the use of innovations in other fields, including but not limited to remote sensing and/or artificial intelligence technology." This proposal is focused on regional/global scale ground failure models that take advantage of both remote sensing and local data using artificial intelligence methods.

**Report:**

**Overview:**
As part of this project, we conducted three independent projects that demonstrated innovative data-driven methods for geospatial ground failure models.

1. An uncertainty quantification framework for logistic regression based geospatial natural hazard modeling (using a geospatial liquefaction model (Zhu et al.2017) for demonstration and implementation.)
2. Regionally informed global geospatial modeling of earthquake-induced soil liquefaction using a system of voting machine learning classifiers building on an updated liquefaction inventory presented in Zhan et al. (2023).
3. Pixel-based Classification Method for Earthquake-Induced Landslide Mapping Using Imagery, Geospatial Data and Temporal Change Information

Each of these projects will be described in brief and is associated with a peer-review publication.

## Section 1. Uncertainty quantification framework for logistic regression based geospatial natural hazard modeling

(The following content has been excerpted from a peer reviewed publication with the same title and the following authors)

Weiwei Zhan, Laurie G. Baise, Babak Moaveni

**Abstract**

There is a class of data-driven global natural hazard predictive models that take advantage of broadly available geospatial proxies. These data-driven geospatial models have been commonly used for landslides and are becoming more available in recent years for liquefaction. Logistic regression is the most common method for predicting these ground failure occurrences. These models do not often include robust quantification of uncertainties although they are widely used in the pre-disaster planning and post-disaster response around the world. Taking the logistic regression based global geospatial liquefaction model (GGLM) (Zhu et al., 2017) as an example, we propose an uncertainty quantification (UQ) framework that consists of characterization of different sources of uncertainty, model sensitivity analysis, and forward uncertainty propagation. In this study, we have identified the main sources of uncertainty in such predictive models as parameter estimation uncertainty, modeling error, and geospatial input uncertainty. A Bayesian inference algorithm is used to quantify the posterior distribution of model parameters and quantify model parameter estimation uncertainties which are found to be negligible when a large amount of data is used in the parameter estimation process. Modeling errors are characterized based on the observed residuals between model predictions and measurements and by fitting a Gaussian distribution to the liquefaction probability residuals. The geospatial input uncertainties are characterized using the literature and expert judgement and propagated into model output. Second, we investigate the sensitivity of model output to different uncertain inputs and find that the variance of model output is largely controlled by the geospatial input uncertainties and model errors. Last, we propose an approximate forward uncertainty propagation method, which provides comparable results to a Monte Carlo simulation-based method with better computational efficiency. The proposed UQ framework provides a measure of uncertainty on model predictions and can be applied to any logistic-regression models and other geospatial modeling problems.

### Uncertainty Quantification for GGLM

Geospatial natural hazard models intrinsically involve different types of uncertainties. In this framework, we identify three types of uncertainty sources: parameter estimation uncertainty, modeling errors, and geospatial input uncertainty. In this section, we quantify (1) estimation uncertainty of model parameters through a Bayesian inference framework, (2) modeling error by fitting a distribution to the observed model prediction residuals, and (3) uncertainty of geospatial explanatory variables from past studies or engineering judgment.

### Bayesian Inference for Quantifying Parameter Estimation Uncertainty

Under the UQ framework, we treat model parameters as random variables and use the Bayes' theorem to estimate the posterior probability distribution of model parameters.

$$p(\boldsymbol{w}|\boldsymbol{D}) = \frac{p(\boldsymbol{D}|\boldsymbol{w})p(\boldsymbol{w})}{p(\boldsymbol{D})} \propto p(\boldsymbol{D}|\boldsymbol{w})p(\boldsymbol{w}) \tag{1}$$

where $\boldsymbol{w}$ is the vector of model parameters to be estimated; $\boldsymbol{D}$ is the vector of available data, i.e., paired geospatial inputs $\boldsymbol{X}$ and liquefaction observation $\boldsymbol{y}$; $p(\boldsymbol{w}|\boldsymbol{D})$ is the posterior distribution of model parameters; $p(\boldsymbol{D}|\boldsymbol{w})$ is the likelihood function as shown in Equation (2); $p(\boldsymbol{w})$ is the prior probability distribution of the model parameters; and $p(\boldsymbol{D}) = \int p(\boldsymbol{D}|\boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w}$ is the evidence which is a normalization constant so the probability of the parameters sums to one.

$$p(\boldsymbol{D}|\boldsymbol{w}) = \prod_{n=1}^{N} p_n^{y_n}(1-p_n)^{1-y_n} \tag{2}$$

In Equation (2), $N$ is the sample size; $n$ is the sample id ranging from 1 to N; $p_n = \frac{1}{1+e^{-Z(x)}}$ is the liquefaction probability for sample $n$ predicted using the logistic function; and $y_n$ is the liquefaction observation for sample n ($y_n = 1$ for liquefaction and $y_n = 0$ for nonliquefaction.
As the solution of posterior distribution of model parameters for logistic regression is intractable (Bishop and Nasrabadi, 2006), we use the Laplace approximation method to estimate $p(\boldsymbol{w}|\boldsymbol{D})$. Laplace approximation aims to find a Gaussian approximation to the posterior probability density defined over a set of continuous variables (Bishop and Nasrabadi, 2006). We assume a Gaussian distribution for the model parameter prior with the general form as shown in Equation (3).

$$p(\boldsymbol{w}) = N(\boldsymbol{w}|\boldsymbol{m_0}, \boldsymbol{S_0}) \tag{3}$$

where $\boldsymbol{m_0}$ and $\boldsymbol{S_0}$ are the mean vector and covariance matrix of the model parameters, respectively. Taking the log of both sides in Equation (1), and substituting Equation (2) for the likelihood function, and Equation (3) for the prior distribution, we obtain the posterior distribution (Bishop and Nasrabadi, 2006):

$$\begin{aligned} ln\, p\,(\boldsymbol{w}|\boldsymbol{D}) = &-\frac{1}{2}(\boldsymbol{w}-\boldsymbol{m_0})^T\boldsymbol{S_0^{-1}}(\boldsymbol{w}-\boldsymbol{m_0}) \\ &+\sum_{n=1}^{N}\{y_n\, ln\, p_n + (1-y_n)\, ln(1-p_n)\} + \text{constant} \end{aligned} \tag{4}$$

To obtain a Gaussian approximation for the posterior distribution, we first find the maximum-a-posteriori (MAP) estimate $\boldsymbol{w}_{\text{MAP}}$ as the maximum of Equation 4, which defines the mean of the posterior distribution. The covariance matrix is then estimated as the inverse of the Hessian (second derivative) of the negative log likelihood, which takes the form

$$\boldsymbol{S_N} = -\nabla\nabla\, ln\, p\,(\boldsymbol{w}|\boldsymbol{D}) = \boldsymbol{S_0^{-1}} + \sum_{n=1}^{N} p_n(1-p_n)\boldsymbol{x}_n\boldsymbol{x}_n^T \tag{5}$$

The Gaussian approximation to the posterior distribution therefore takes the form

$$q(\boldsymbol{w}) = N(\boldsymbol{w}|\boldsymbol{w_{MAP}}, \boldsymbol{S_N}) \tag{6}$$

The prior distribution uses zero means for all the model parameters, and standard deviation of 100. In that way, the prior distribution is flat, which indicates an uninformative prior distribution. The Laplace approximation is then used to estimate the joint posterior distribution of model parameters.

The pairwise correlations between model parameters are estimated using the $\boldsymbol{S_N}$ by converting the covariance matrix to correlation matrix.

The mean, standard deviation, and coefficient of variation (COV) of the GGLM model parameters are summarized in Table 1. The model parameters show small estimation uncertainty, with the maximum absolute value of COV being 2.6% (for ln(*precip*) and ln(*wtd*)), not exceeding 1.0% for the remaining model parameters. The pairwise correlations between the model parameters of GGLM are shown in Table 2. Two model parameters (ln($V_{S30}$) and ln(*wtd*)) have strong correlations with intercept, and ln($V_{S30}$) also shows a strong negative correlation with ln(*wtd*). The positive parameter correlation means the model parameter tends to increase with the increase of another model parameter, and vice versa (Li and Vu, 2013). Statistically, the maximum likelihood fitting (Equation 2) cannot guarantee unique parameter estimation, due to correlations among the parameters. The correlation can be explained by the physical background of liquefaction phenomena. For instance, the site stiffness ($V_{S30}$) and water table depth (*wtd*) can affect the soil liquefaction susceptibility jointly. In short, the Bayesian inference results suggest the GGLM tend to have low parameter estimation uncertainty.

**Table 1.** Parameter estimation uncertainty of the geospatial liquefaction hazard model.

| Coefficients | Intercept | $w_{\ln(\mathrm{PGV})}$ | $w_{\ln(VS30)}$ | $w_{\ln(precip)}$ | $w_{\ln(dw)}$ | $w_{\ln(wtd)}$ |
|---|---|---|---|---|---|---|
| Mean | 6.731 | 0.279 | -1.459 | 0.167 | -0.695 | -0.130 |
| Std | 0.067 | 0.003 | 0.013 | 0.004 | 0.004 | 0.003 |
| COV (%) | 1.0 | 1.0 | -0.9 | 2.6 | -0.6 | -2.6 |

**Table 2.** Correlation matrix for the model parameters of the geospatial liquefaction hazard model.

| Correlation | Intercept | $w_{\ln(\mathrm{PGV})}$ | $w_{\ln(VS30)}$ | $w_{\ln(precip)}$ | $w_{\ln(dw)}$ | $w_{\ln(wtd)}$ |
|---|---|---|---|---|---|---|
| Intercept | 1.00 | -0.02 | **-0.88** | -0.24 | 0.11 | **0.70** |
| $w_{\ln(\mathrm{PGV})}$ | | 1.00 | -0.14 | 0.08 | -0.06 | 0.09 |
| $w_{\ln(VS30)}$ | | | 1.00 | -0.23 | -0.14 | **-0.80** |
| $w_{\ln(precip)}$ | | | | 1.00 | -0.01 | 0.16 |
| $w_{\ln(dw)}$ | | | | | 1.00 | 0.06 |
| $w_{\ln(wtd)}$ | | | | | | 1.00 |

## Residual Analyses for Quantifying Modeling Error

### Modeling Error

Modeling errors are due to imperfections of the simplified statistical models in representing complex natural hazard phenomena and inherent randomness of geophysical processes. For geospatial natural hazard models with continuous outputs (i.e., regression models), such as earthquake ground-motion models (Boore et al., 2014), the modeling errors can be defined as the differences between the observed and predicted hazard intensities. However, for geospatial models with categorical target variables (i.e., classification models), the modeling errors are not easily

defined (Liu and Zhang, 2018). Here, we quantify the modeling errors of GGLM in the space of liquefaction probability. Specifically, the model residuals of the GGLM are defined as the differences of empirical and theoretical liquefaction probabilities which are both conditional on the $Z$ value.

$$\epsilon(Z) = P(Z)_{empi} - P(Z)_{theo} \quad (7)$$

where $\epsilon$ is the liquefaction probability residual which is considered as modeling error; $P(Z)_{empi}$ is the empirical (observed) liquefaction probability which is computed as the ratio of the liquefaction sample size to the total sample size for each Z value bin; and $P(Z)_{theo}$ is the theoretical (predicted) liquefaction probability.

## Global Residual Model

The relationship between the empirical and theoretical liquefaction probabilities is shown in Figure 1. The empirical liquefaction probabilities are represented for different Z bins based on the corresponding proxies of observed locations, and they match well with the theoretical ones for samples with $Z \in [-2.2, 1.4]$ (i.e., $P(Z)_{theo} \in [0.1, 0.8]$). It is noted that the observed liquefaction probabilities deviate from the predicted liquefaction probabilities at the low and high Z values, which is likely due to insufficient sample size at these bins for calculating empirical liquefaction probabilities. We use 150 as the threshold sample size per bin to exclude outlier bins based on the relationship between liquefaction probability residuals and the sample size per bin (Figure 1b).

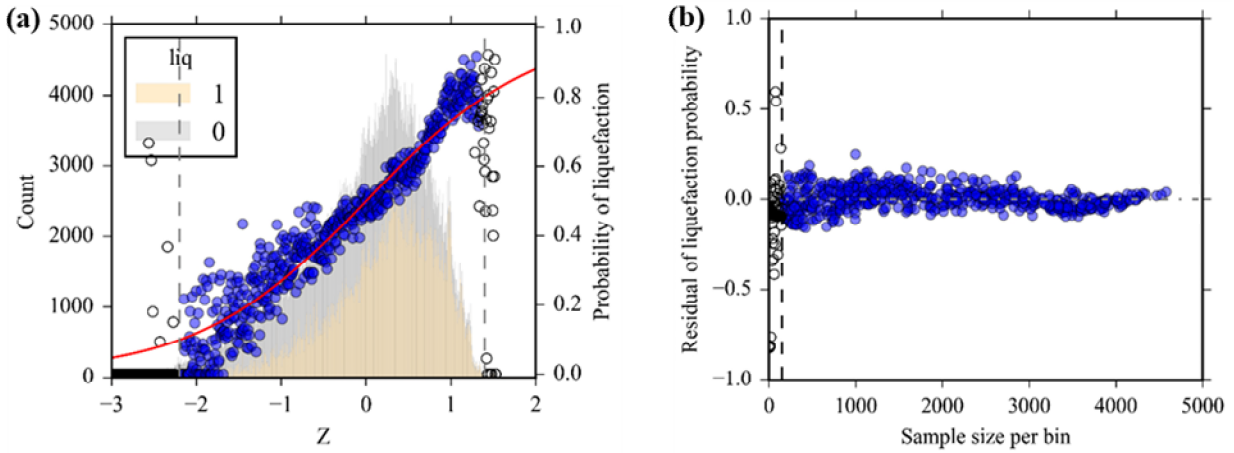


**Figure 1.** (a) Comparison between empirical (circles) and theoretical (red curve) liquefaction probabilities. The empty circles indicate bins with sample size less than 150. The histogram shows the number of liquefaction and non-liquefaction samples. The vertical dashed lines constrain the data range with sufficient sample size for computing empirical liquefaction probability. (b) The correlation between liquefaction probability residuals and sample size per bin. The vertical dashed line indicates sample size for reliable estimation of empirical liquefaction probability.

The liquefaction probability residuals show different patterns at varying Z values. The GGLM consistently overestimates the liquefaction probability at the very low Z value range (Z< -2.2) where the empirical liquefaction probabilities are zero except for a few abnormal high values due to insufficient sample size. At the very high Z values (Z>1.4), the model also tends to have a smaller number of reliable estimations of empirical liquefaction probability. For the middle $Z$ value range ($Z \in [-2.2, 1.4]$), the liquefaction probability residuals have slightly higher variability for

negative Z values than for positive values but generally follow a Gaussian distribution as shown in Figure 2b. In this study, we represent the liquefaction probability residuals as a Gaussian distribution with mean of zero (as the fitting process is unbiased) and standard deviation of 0.06 (2b). Based on the residual analyses, we propose conditional mean and residual models for global application of the GGLM (Table ) which is basically a truncated Gaussian distribution model. For locations with Z between -2.2 to +1.4, mean estimate of liquefaction probability is the direct output of the logistic regression model while modeling error is represented by the fitted Gaussian distribution in Figure 2b. For locations with Z<-2.2, we recompute the empirical liquefaction probability as the ratio of the number of the liquefaction samples with Z <-2.2 to the total number of samples with Z <-2.2, and assign this empirical probability value (i.e., 0.05) as the mean estimate of liquefaction probability with Z<-2.2. For locations with Z>1.4, we assign the mean estimate of liquefaction probability as 0.8 using the similar method but removing the samples from several outlier Z bins. We assume the residual models for the locations with insufficient samples (i.e., Z<-2.2 and Z>1.4) fit the same Gaussian distribution calibrated using the main part of dataset (2b) although we do not have enough reliable data to calibrate them. It is noted that our residual analyses also contribute to cap the predicted liquefaction probability.

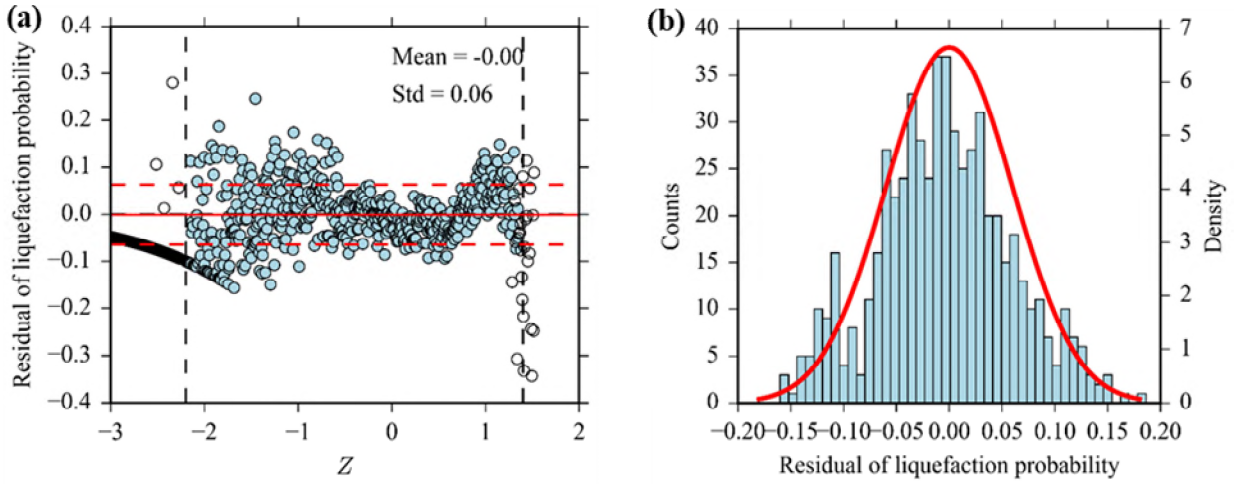


**Figure 2.** (a) Relationship between the liquefaction probability residuals and the Z value. The circles with no fill have insufficient samples to estimate empirical liquefaction probability (i.e., the sample size is less than 150). (b) Histogram of the liquefaction probability residuals. The red line is the theoretical probability density function for the Gaussian distribution with mean of zero and standard deviation of 0.06.

**Table 3.** Proposed mean and modeling error forms of the considered GGLM after the residual analyses.

| Stage | Applicable range | $P(y=1)$ | $\epsilon$ |
|---|---|---|---|
| 1 | $Z$ < -2.2 | 0.05 | Assumed *N(0, 0.06)** |
| 2 | -2.2≤$Z$≤1.4 | $\frac{1}{1+e^{-Z(x)}}$ | *N(0, 0.06)* |
| 3 | Z > 1.4 | 0.80 | Assumed *N(0, 0.06)* |

* $N(\mu,\sigma)$ denotes the Gaussian distribution where μ is the mean and σ is the standard deviation.

## Section 2. Regionally informed global geospatial modeling of earthquake-induced soil liquefaction using a system of voting machine learning classifiers.

(The following content has been excerpted from a peer reviewed publication with the same title and the following authors)

Adel Asadi[1], Laurie Gaskins Baise[1], Snehamoy Chatterjee[2], Weiwei Zhan[1,3], Alexander Chansky[1], Babak Moaveni[1]

[1] Geohazards Research Lab, Department of Civil and Environmental Engineering, School of Engineering, Tufts University, Medford, MA 02155, USA.

[2] Geological and Mining Engineering and Sciences Department, Michigan Technological University, Houghton, MI 49931, USA.

[3] Texas Advanced Computing Center, Civil, Architectural and Environmental Engineering Department, The University of Texas at Austin, TX 78758, USA.

**Abstract:**

Data-driven geospatial liquefaction models are useful tools for real-time post-event impact and regional seismic hazard assessments. Geospatial liquefaction models are based on liquefaction occurrence inventories, widely available geospatial variables, and earthquake-specific parameters. This research uses an updated inventory with geospatial data from non-liquefaction and liquefaction occurrence locations in 53 earthquakes around the world, including categorical and continuous variables representing proxies for soil saturation, soil density, and earthquake loading. In this study, the performance of advanced machine learning (ML) algorithms in learning complex nonlinear patterns in the large dataset to predict liquefaction susceptibility, as an alternative approach to previously published logistic regression classifiers is evaluated. The proposed methodology starts with an exploratory data analysis and feature selection scheme on the sampled data across both the liquefaction and non-liquefaction classes to remove redundant and less relevant features, run data transformations, and perform statistical analysis to define some classification thresholds. The class and event imbalance issues are treated in an innovative manner by under-sampling large events and distributing the datasets over several balanced subsets. Instead of training a single classifier, a system of voting machine learning classifiers is designed to achieve higher accuracy and certainty, using different classification techniques to balance the performance, as not a single classification technique can be considered the best for all earthquake events. The voting classifiers use different portions of the data, representing global, coastal/non-coastal and regional information, to optimize the predictive power of the proposed model by taking the majority votes of six different classifiers. The classifiers include three global classifiers, two coastal/non-coastal classifiers, and a regional classifier. For each data subset, the ML-based binary classification model is trained and validated via a K-fold cross-validation approach, and the voting system is tested based on the leave-one-out approach, excluding individual earthquake events one at a time for a fair accuracy, reliability, and applicability assessment. The final class assignment is generated by majority vote of the system of 6 voting classifiers trained by the individual data subsets. The results of this study are compared with a model developed using logistic regression to investigate the benefits and limitations of the proposed voting approach.

### Methodology

In this research, a novel system of voting machine learning classifiers is proposed as an alternative to logistic regression based geospatial liquefaction models. The system of voting machine classifiers allows for the inclusion of multiple global models, coastal and noncoastal models, and geographic regional models. The dataset is balanced across events by considering a cap for the number of samples per class. The sampled imbalanced data is distributed into subsets of balanced training data: 3 global models, 2 coastal models, 2 non-coastal models, and 6 geographic regional models. The system of classifiers is implemented through both majority voting and soft voting of the classifiers. The model performance is assessed with a leave-one-out testing approach, which means that data of individual earthquake events are excluded from the training data in a loop over all events. The flowchart of the proposed system is presented in Figure 3.

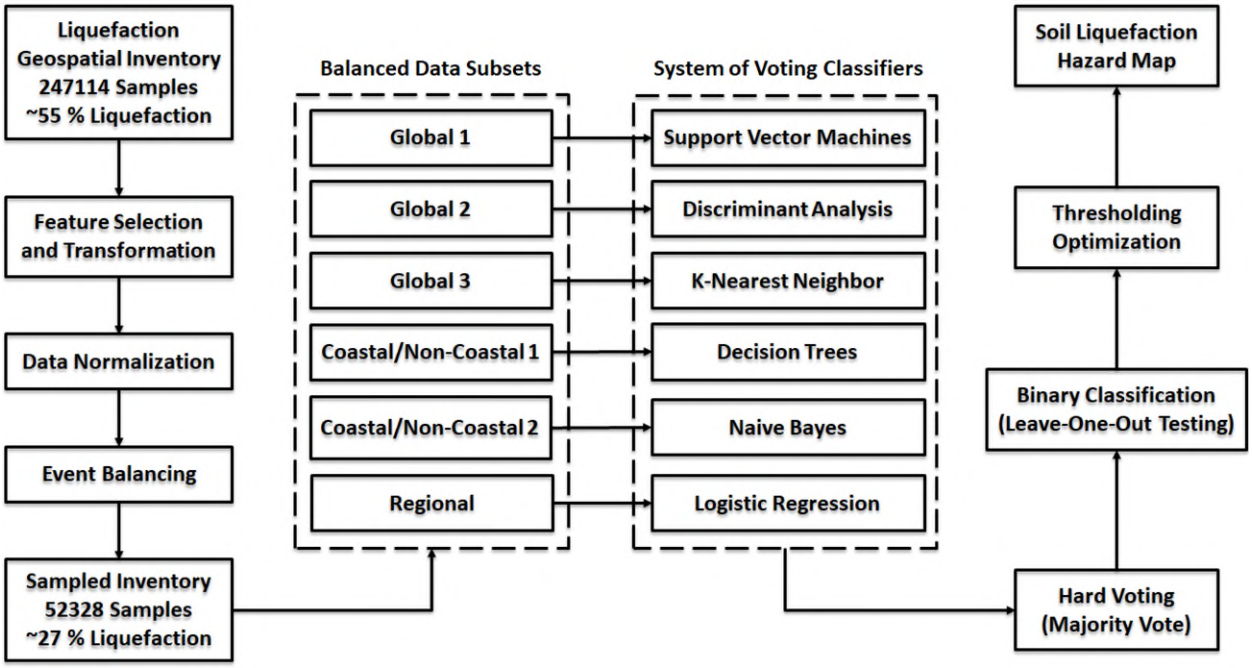


**Figure 3.** Flowchart of the proposed voting classification approach. Since 6 regions are considered in the inventory, 6 different regional models will be utilized based on the individual earthquake being tested.

## Exploratory Data Analysis

Different types of exploratory data analysis (EDA) technique are used to evaluate the features in terms of redundancy and relevance to the binary classification problem in this research. A variety of univariate (histograms, box plots, bar plots, etc.) and bi-variate (correlation analysis, etc.) analysis methods were considered to evaluate the parameters' class separability and their relationship with each other. Feature transformations were proposed to improve the performance of several features in distinguishing the classes of liquefaction and non-liquefaction. After the EDA process, the data of all features ($X_i$), except the categorical variables are normalized via *Z*-score method to have a mean of zero and standard deviation of one, as presented via Equation 8, where $\mu$ is the mean, and $\sigma$ is the standard deviation of each variable. Most machine learning algorithms perform better when variables have a Gaussian distribution (Geyin et al., 2022), all predictors were transformed and normalized.

$$Z = \frac{X-\mu}{\sigma} \quad (8)$$

**Data Imbalance**

Data imbalance can be problematic, as machine learning algorithms for classification problems are designed for balanced datasets; otherwise, their predictions will be biased toward the majority classes, as they were less punished for misclassifying the minority classes during the training and validation process (Weiss and Provost, 2001). The spatially incomplete nature of the liquefaction inventory database and the associated class imbalance has a strong influence on the probabilities of the model (Oommen et al., 2011; Thompson et al., 2016-b; Zhu et al., 2015 & 2017). One solution to this issue is to under-sample the majority class by random sampling (Fernandez et al., 2019). The other solution could be to over-sample the minority class using synthetic data generation techniques. Bozzoni et al. (2020 & 2021) resolved the issue of class imbalance by trying commonly used synthetic data generation (over-sampling) algorithms, including SMOTE (Chawla et al. 2002) and ADASYN (He et al., 2008). However, such methods can be uncertain in terms of losing valuable information in under-sampling and adding noise and unrealistic samples to the data in over-sampling process. Zhu et al. (2015) chose to use a highly imbalanced dataset (almost 13 non-liquefaction samples per each liquefaction record) because they aimed to develop a probability estimator that predicts the areal extent of liquefaction and wanted the probability to correlate to the spatial extent. However, in their later version, Zhu et al. (2017) balanced their database after performing the non-liquefaction sampling process. Similarly, Baise et al. (2021) used a balanced sampling strategy, where the more significant liquefaction events are under-sampled so that they don't dominate the model. They capped the number of liquefaction and non-liquefaction points from each earthquake to 2000 and 1000 points, respectively.

In this study used an alternative approach. The dataset is balanced across the events by considering a cap for the number of samples per class (2085 for non-liquefaction and 2577 for liquefaction). Then, the resulting dataset which contains ~27 % liquefaction samples are distributed over six various balanced subsets, including three sets of global data, two sets of coastal/non-coastal data (depending on the earthquake being tested), and a set of six geographical regional datasets, based on the regional categorization of the events. Because the method is demonstrated with the leave-one-out approach to model evaluation, the data of the tested earthquake is not included in any of the data subsets.

**Feature Selection via Machine Learning**

After removing a few features (variables) through the EDA process, a feature selection scheme is implemented to remove any features with zero or near-zero weight in the classification process. The strategy in this research is to keep explanatory variables as long as they are informative. The goal of feature selection is to reduce the risk of model over-fitting, to improve accuracy, and reduce training time. Neighborhood component analysis (NCA) is used in this study as the feature selection technique. NCA is a machine learning algorithm for metric learning that learns in a supervised manner for improving the classification accuracy of a stochastic nearest neighbors' rule in the transformed space (Goldberger et al. 2005). The algorithm works best to estimate feature importance for any distance-based supervised models that use pair-wise distances between observations to predict the response (Guyon and Elisseeff, 2003). NCA feature selection is performed with regularization to learn feature weights for minimization of an objective function that measures the average leave-one-out classification loss over the training data (Yang et al.,

2012). The regularization term (λ) in NCA's objective loss function derives the weights of irrelevant predictors to zero. As the number of observations n increases, the chance of overfitting decreases, and the required amount of regularization also decreases. λ=1/n, where, n is the number of samples, is an approximately good selection for the regularization parameter. Stochastic gradient descent algorithm was used as the solver to estimate feature weights.

## Classification Modeling

The logistic regression models developed in the studies by Zhu et al. (2015), Zhu et al. (2017), and Baise et al. (2021), were based on an intentional strategy to develop a simple model that uses parameters that reflect the underlying physics of the problem and can be used easily and quickly for future events. In this work, the goal is to increase the predictive accuracy through the use of a complex machine learning model.

## Classification Thresholds

Thresholds have commonly been used in geospatial liquefaction models to prevent overfitting and to include thresholds related to physical limits to liquefaction occurrence. The GGLM model (Zhu et al., 2017) used a PGV minimum threshold of 3 cm/s. Rashidian and Baise (2020) modified the GGLM model by recommending a PGA minimum threshold of 0.1 g as an additional threshold to better characterize the step-function behavior of the PGA plot. A threshold of 0.1 g for horizontal peak ground acceleration (PGA) was also used by Bozzoni et al. (2020) to reduce the spatial extent of liquefaction from prior geospatial liquefaction models, based on experimental data showing that liquefaction triggering is highly unlikely for PGA values of less than 0.10 g (Green and Bommer 2019). Based on the inventory used herein, liquefaction has occurred in the 2012 M6 Emilia Earthquake in Italy when PGA < 0.1 g indicating that this threshold may be too high.

The US Geological Survey (USGS) has also implemented a magnitude scaling factor (MSF) to the PGV term in the GGLM equation to reduce the over-prediction for low-magnitude earthquakes (M < 6). Rashidian and Baise (2020) have left the decision on using MSF to the GGLM model users, since they didn't observe any considerable positive or negative impact in their comparative analysis for high-magnitude earthquakes where over-prediction is more likely to happen, although they confirmed that the spatial extent of liquefaction will be moderated for low-magnitude events when using this factor.

Zhu et al. (2017) also implemented a maximum threshold of 620 m/s for the $Vs_{30}$, which means that any point with a $Vs_{30}$ value of greater than the threshold is automatically considered as non-liquefaction. In the current inventory, this Vs30 threshold leads to under-prediction in for the 1993 M7.7 Hokkaido earthquake in Japan. Rashidian and Baise (2020) also recommended the use of an upper bound for GGLM (Zhu et al., 2017), when annual precipitation of a region exceeds 1700 mm, which was the upper quartile of annual precipitation in their development database of earthquakes, since overprediction of liquefaction was observed in regions with high precipitation. The annual precipitation threshold reduced the over-prediction of liquefaction in the 14 affected earthquake events in their study.

The use of thresholds has also been used in the study by Jena et al. (2022) in their deep learning model of liquefaction hazard for Australia. They have applied thresholds for different input variables including magnitude of Mw6. Green and Bommer (2019) concluded that earthquakes as small as Mw 4.5 can trigger liquefaction in extremely susceptible soil deposits, based on their field observations and a simple parametric study. However, they have indicated that

for soil profiles which are appropriate for building structures, the minimum Mw 5 can be considered for liquefaction triggering (Green and Bommer, 2019). In this study, the threshold value of Mw 5 is considered as the minimum liquefaction triggering earthquake magnitude, although the lowest magnitude in the inventory, leading to liquefaction occurrence was 5.8 Mw.

In the current version of the GGLM models (GGLM-21-a and GGLM-21-b), Baise et al. (2021) do not retain the thresholds established by Zhu et al. (2017) and Rashidian and Baise (2020) around PGA, PGV, precipitation, and $Vs_{30}$, since only 1.5% of the liquified data points in their inventory violate those thresholds. In the current study, we use the magnitude judgement-based threshold and then rely on data-driven thresholds. Data driven thresholds are only used for the features that are used in the final model, and do not violate the physics-based criteria of liquefaction triggering. It should be noted that the thresholds are designed to optimize the model performance as a post-processing step after the classification modeling, and no sample bypasses the classification step.

## System of Voting Classifiers

This study proposes a system of voting machine learning classifiers to create liquefaction hazard maps. There are several benefits in implementing such approach, including 1) Reduction in sensitivity to under-sampling the dataset for the purpose of balancing, and taking advantage of all available valuable data, instead of under-sampling the majority class; 2) Reduction in model errors/uncertainties and inaccurate predictions by taking average of several model predictions, leading to stabilization of predictions and better generalizability and consistency of hazard modeling; 3) Benefiting from different type of classifiers, since not a single classifier is the best to predict all events; 4) Regional and coastal/non-coastal optimization of the classification decision by including these special classifiers in addition to the global classifiers; and 5) The ability to generate probability maps in a continuous form (soft voting) rather than 0-1 binary (hard voting) prediction.

Six different algorithms are used in this study, as shown in the model flowchart via Figure 5. These algorithms are widely used machine learning algorithms for classification purposes. They include support vector machines (SVM), discriminant analysis, K-nearest neighbor (KNN), decision trees, Naïve Bayes, and logistic regression, which are used as three global, two coastal/non-coastal, and a regional classifier, respectively. The k-fold cross-validation was used within the training step for all 6 classifiers to avoid overfitting (Kohavi, 1995). The method divides the training dataset into k (k=5 in this study) smaller folds with equal number of samples points for the optimization of parameters. The hyperparameters and the architecture of the proposed classification algorithms were optimized separately with a grid search approach during the training process to achieve the best performance. A brief description of the classification algorithms used in this study are provided in the following sub-sections.

## Support Vector Machines

Support vector machines are widely used and reported among the best performing binary classifiers, although they are also used to solve multi-class problems. The objective of the SVM algorithm is to find a hyperplane that, to the best degree possible, separates data points of one class from those of another class with the largest margin between the two classes. Margin means the maximal width of the slab parallel to the hyperplane that has no interior data points. For linearly separable problems, the algorithm can find the hyperplane; however, for most practical problems, the algorithm maximizes the soft margin allowing a small number of misclassifications. Support

vectors refer to a subset of the training observations that identify the location of the separating hyperplane (Vapnik, 1995; Christianini and Shawe-Taylor, 2000; Hastie et al., 2009). In this study the efficient linear SVM is used as a classifier for one of the global models in the voting classification system. In this study, the efficient linear SVM is used as a classifier for one of the global models in the voting classification system, which is less prone to over-fitting.

**Discriminant Analysis**

Discriminant analysis is used as one of the supervised classification methods, which assumes that different classes generate data based on different Gaussian distributions (Fisher, 1936). To train the classifier, the fitting function estimates the parameters of a Gaussian distribution for each class and finds the smallest calculated misclassification cost to predict classes of new (unlabeled) data. The linear discriminant analysis (LDA) used in this research, computes the sample mean of each class. Then it computes the sample covariance by first subtracting the sample mean of each class from the observations of that class and taking the empirical covariance matrix of the result (Guo et al., 2007).

**K-Nearest Neighbors**

The K-nearest neighbors' algorithm is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. Categorizing query points based on their distance to points in a training data set can be a simple yet effective way of classifying new points. K-nearest neighbor (KNN) classifiers can use various metrics (Euclidean distance, in this study) to determine the distance between a set of data and query points, and simply ranks the points by their distance to the query points in order to classify them as the closest class, assuming that similar points can be found near one another. A class label is assigned based on a majority vote, meaning the label that is most frequently represented around a given data point is used (Cover and Hart, 1967). The value of K is optimized to avoid over-fitting and under-fitting on the global data.

**Decision Trees**

Decision trees, or supervised non-parametric classification trees, predict responses to data inputs, by following the decisions in the tree from the root (beginning) node down to a leaf (end) node, which contains the response (Breiman et al., 1984; Rokach and Maimon, 2008). In tree structures, which are famous for their simplicity and easy interpretability, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Classification trees give responses that are nominal, such as liquefaction or non-liquefaction. A tree is built by splitting the source set, constituting the root node of the tree, into subsets which constitute the successor children. The splitting is based on a set of splitting rules based on classification features (Shalev-Shwartz and Ben-David, 2014). This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node has all the same values of the target variable, or when splitting no longer adds value to the predictions. The cost of using the tree (i.e., predicting data) is logarithmic in the number of data points used to train the tree. Thus, the classifier is assigned to the coastal/non-coastal part of the model, which can deal with low number of data when the tested event is non-coastal.

**Naïve Bayes**

Naïve Bayes is a classification algorithm that applies density estimation to the data. The naïve Bayes classifier is designed based on the naïve assumption that predictors are independent of one another within each class (Hastie et al., 2009). While the class-conditional independence between predictors is not true in general, research shows that this optimistic assumption works well in practice. It classifies data in two steps: 1) Using the training data, the method estimates the parameters of a probability distribution, assuming predictors are conditionally independent given the class; then, 2) For any unseen test data, the method computes the posterior probability of that sample belonging to each class. The method then classifies the test data according to the largest posterior probability (Manning et al., 2008). In this study, normal (Gaussian) predictor conditional distribution is used in the training function. This assumption of class-conditional independence of the predictors allows the naive Bayes classifier to estimate the parameters required for accurate classification while using less training data than many other classifiers. That is the reason that Naïve Bayes classifier is assigned to the coastal/non-coastal part of the model, which can deal with low number of data when the tested event is non-coastal. This makes it particularly effective for data sets containing many predictors.

### Logistic Regression

Logistic Regression is a supervised machine learning algorithms which is commonly used for binary class classification problems. Logistic regression was borrowed by machine learning from the field of statistical models and is currently one of the most well-known and widely used machine learning algorithms (James et al. 2013). Logistic regression operates using a sigmoidal function for values ranging between 0 and 1. A logistic regression algorithm takes as its input a feature vector and outputs a probability, where the feature vector represents an object belonging to the positive class. Since the logistic regression model is very simple, compared to the other classifiers used in this study, it is assigned to the regional datasets, as these datasets are expected to have more common characteristics, compared to global and coastal/non-coastal models.

### Model Validation

### Leave-One-Out Testing

The validation of the system is done by excluding data from an individual earthquake from the training process, and independently evaluating the inventory for that event in a leave-one-out testing approach. This process is done to fairly assess the performance of the model for new unseen earthquakes. This approach is different than the validation scheme followed by previous studies (Zhu et al., 2017; Todorovic and Silva, 2022), where they use all or a large portion of samples from all earthquakes in the training and model development process. This approach leads to a better understanding of the performance of the system in an unbiased way and provides insights into system of classifiers.

Figure 4 provides a visual explanation of the model assessment and implementation and provides an example of how the soft and hard voting systems work with data from an unseen earthquake event. In this study, hard voting considering the majority votes of the classifiers. If at least 3 classifiers label a sample with liquefaction (class 1), the point will be classified as liquefaction. As an alternative, the soft voting can generate probability-based liquefaction maps that take into account epistemic uncertainty. The soft voting decision is made by taking the average of the class labels predicted by the classifiers in the voting system. In this research, since multiple classifiers are used to automatically balance each other's outputs, the majority hard voting (3+

votes in favor of either class) is considered as the suitable boundary between the two classes and is similar to the optimal threshold used by Bozzoni et al. (2020) in their logistic regression model (0.57).

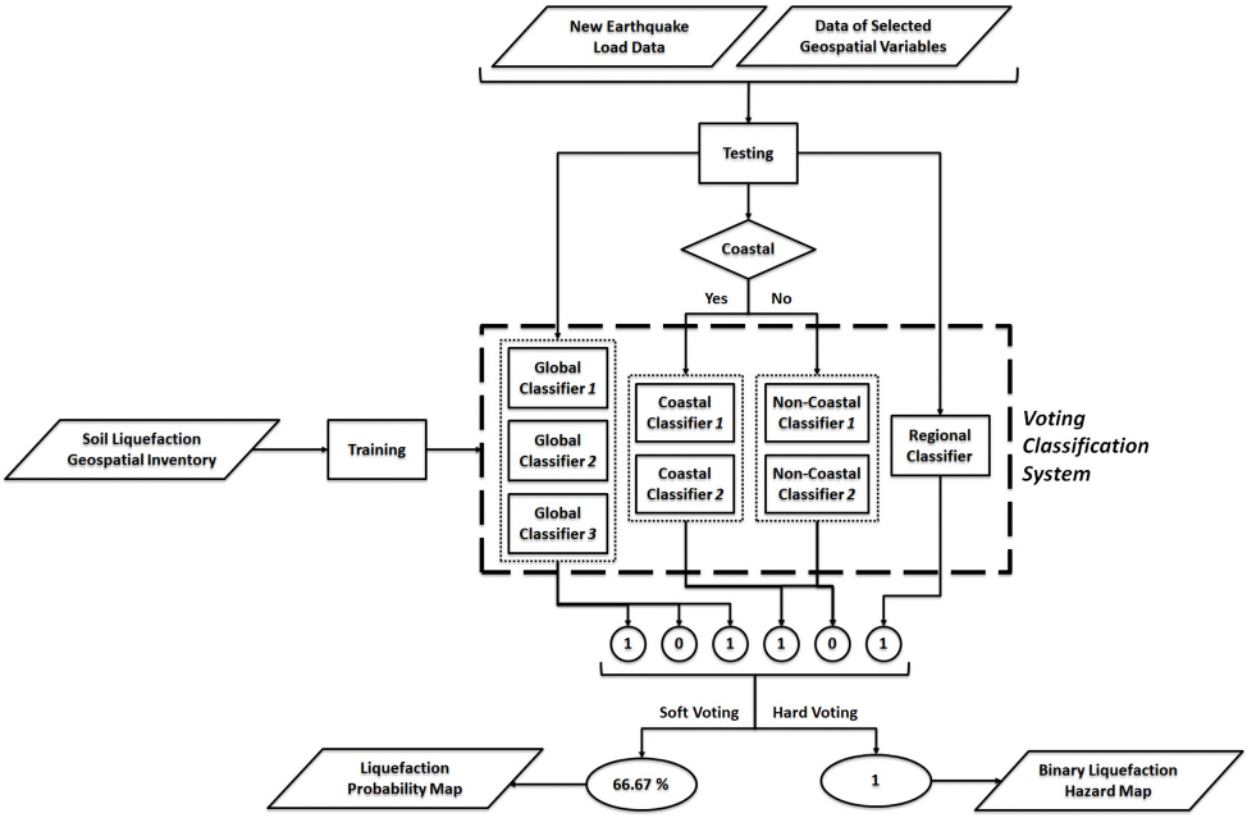


**Figure 4.** The leave-one-out testing approach used in this study, plus an example of how the final classification decision is made via hard voting or soft voting procedures. Global classifiers 1, 2 and 3 are SVM, discriminant analysis, and KNN, respectively. Coastal/non-coastal classifiers 1 and 2 are decision trees and Naïve Bayes, respectively, and the regional classifier uses logistic regression.

## Accuracy Indices

In order to evaluate accuracy, we use the hard voting with majority voting so that each pixel is labelled as 1 (liquefaction or positive class), or 0 (non-liquefaction or negative class). The performance of the model was assessed via ROC (Receiver Operating Characteristics) analysis and expressed via the Area Under Curve (AUC) parameter. In a ROC curve, the Sensitivity or True Positive Rate (TPR) is plotted against the Specificity or False Positive Rate (FPR). The TPR and FPR are calculated via Equations 9 and 10.

$$\text{Sensitivity} = TPR = \frac{TP}{TP+FN} \tag{9}$$

$$Specificity = FPR = \frac{FP}{TN+FP} \tag{10}$$

Where, TP (True Positive) is liquefaction predicted as liquefaction, FP (False Positive) is non-liquefaction predicted as liquefaction, TN (True Negative) is non-liquefaction predicted as non-liquefaction, FN (False Negative) is liquefaction predicted as non-liquefaction. The AUC ranges between 0.5 (random classifier) and 1 (100% accurate classifier). These accuracy indices (Sensitivity, Specificity and AUC) are provided for the models alongside the Prevalence, which

shows the percentage of the specific earthquake dataset belonging to the positive class (liquefaction). The AUC values are compared with previous works and among the models presented in this study.

## Results and Discussion

### Exploratory Data Analysis

In the exploratory data analysis (EDA) step, two main goals were pursued: 1) EDA-based and knowledge-based feature removal; and 2) Implementation of appropriate feature transformations. As a result, several features were removed or transformed prior to the machine learning-based feature selection. Table 4 provides the results of feature removal and the reasons behind the decisions.

PGA is more frequently used in liquefaction modeling methods (Seed and Idriss, 1971; Zhu et al., 2015; Bozzoni et al., 2020) because it is proportional to the maximum shear stress induced in the sediment (Terzaghi et al., 1996). However, PGV has shown a stronger correlation with liquefaction occurrence than PGA in several studies (Zhu et al., 2017; Baise et al., 2021 & 2023), implying that soil liquefaction could be more sensitive to the low-frequency components of the ground motion, and liquefaction is more dependent on input kinetic energy, which is well-characterized by the PGV (Kotoda et al., 1986; Midorikawa and Wakamatsu, 1988; Liang et al., 1995; Bardet and Liu, 2009). Additionally, Bardet and Liu (2009) observed nonlinear behavior between the PGV and empirical probability of liquefaction; As PGV increases, the probability of liquefaction first rapidly increases and then reaches a plateau when PGV is greater than a threshold of 10 cm/s. Therefore, PGA was removed from the features based on the previous knowledge in the published literature, and PGV was retained as the loading feature.

**Table 4.** EDA-based and knowledge-based feature removal results.

| Number | Removed Feature | Reasoning |
|---|---|---|
| 1 | DR2 | Repetitive information from alternative source |
| 2 | DC2 | Repetitive information from alternative source |
| 3 | Precipitation | o Correlation with AI (redundancy)<br>o Previous knowledge from literature on over-prediction issues (Rashidian and Baise, 2020) |
| 4 | PGA | o Previous knowledge from the literature on the superiority of PGV in correlating with liquefaction occurrences (Zhu et al., 2017)<br>o Correlation with PGV |
| 5 | DWB2 | Repetitive information from alternative source |
| 6 | HWB2 | Repetitive information from alternative source |
| 7 | Slope | Correlation with TRI (redundancy) |
| 8 | Roughness | Correlation with TRI (redundancy) |
| 9 | Aspect | No observable separability of classes in univariate and bi-variate analysis with other features (irrelevance) |
| 10 | Landform | High number of classes leading to several categorical dummy variables in the classification modeling |

| | | |
|---|---|---|
| 11 | Landcover | Binary categorical data, not desirable based on the designed voting classification method |

Feature transformations were applied when the transformation resulted in increased ability to distinguish the two classes. Table 5 presents the feature transformations implemented in this study. PGV is well-represented by a lognormal distribution; thus, PGV was transformed by taking the natural logarithm. The same transformation was applied to the $Vs_{30}$ values. These transformations have been used by several previous studies (Zhu et al., 2017; Rashidian and Baise, 2020; Bozzoni et al., 2020; Baise et al., 2021; Todorovic and Silva, 2022). The distributions of some variables were skewed (distance to the waterbody and water table depth); thus, they were transformed by taking the square root of the original values. This transformation was also used by Todorovic and Silva (2022). TPI and TRI were transformed in the study by Baise et al. (2021) by taking the square-root of the absolute value. Figure 5 illustrates the benefit of feature transformations on the TPI variable. Baise et al. (2021) have also used square-root transformation for distance to the closest waterbody, elevation above the closest waterbody, and elevation.

**Table 5.** Feature transformation techniques used in this study. "Ln", "SQRT", "ABS", and "x", stand for natural logarithm, square root, absolute value, and the original value, respectively.

| Number | Feature | Transformation Technique |
|---|---|---|
| 1 | PGV | Ln (x) |
| 2 | DL | SQRT (x) |
| 3 | DWB | SQRT (x) |
| 4 | HWB | SQRT (ABS (x)) |
| 5 | WTD | SQRT (x) |
| 6 | HAND | SQRT (x) |
| 7 | $Vs_{30}$ | Ln (x) |
| 8 | TPI | SQRT (ABS (x)) |
| 9 | TRI | SQRT (ABS (x)) |
| 10 | Elevation | SQRT (x) |

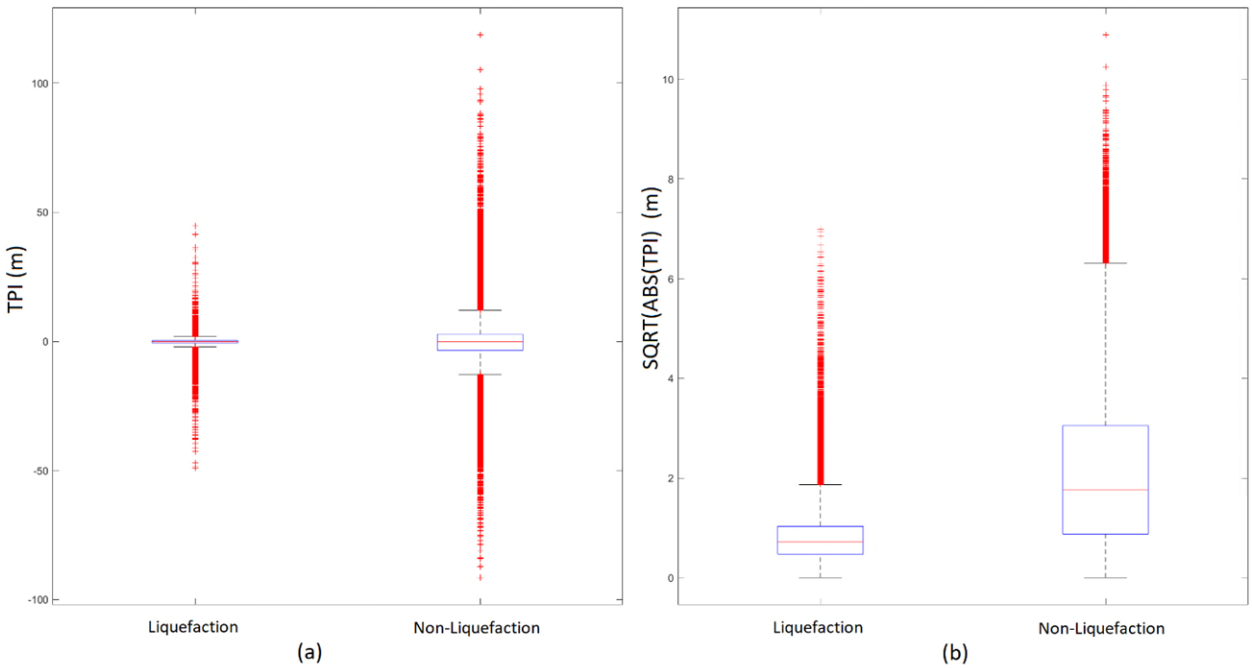


**Figure 5.** a) Comparative box plots of the original TPI data of the two classes, versus b) Transformed TPI data (square root of the absolute value), showing the shift in their comparative means, and better distinguishment between the two classes.

### Feature Selection via Machine Learning

Neighborhood component analysis (NCA) feature selection was used to compare the weight of the features, and to remove the features with zero or near-zero weight. Based on the results provided in Table 6, all the features showed contribution to the classification; therefore, all remaining 17 variables are retained for the classification models. To improve the certainty of NCA performance, and because of the stochastic manner of NCA, the datasets were partitioned into 5 subsets, and the average weight of features on the NCA analysis of the subsets is reported in Table 6, ranked from high to low. It should be noted that the weights are based on the combined use of the variables, as NCA does not evaluate features independently, which is one of the reasons for using this algorithm at this feature selection step. As an example, on the importance of integrated feature selection, Zhu et al. (2017) reported in interaction assessment (bivariate analysis), that distance to river does not appear to be a good predictor when it is evaluated alone, but it becomes valuable when combined with distance to coast.

**Table 6.** NCA feature selection results.

| Rank | Feature | Weight |
|---|---|---|
| 1 | AI | 5.702 |
| 2 | PGV | 5.409 |
| 3 | Magnitude | 5.300 |
| 4 | DL | 4.558 |
| 5 | DC | 4.350 |
| 6 | DWB | 3.413 |
| 7 | Elevation | 3.388 |

| | | |
|---|---|---|
| 8 | $V_{S30}$ | 3.273 |
| 9 | DR | 3.254 |
| 10 | HWB | 2.966 |
| 11 | HAND | 2.961 |
| 12 | Thickness | 2.452 |
| 13 | TRI | 2.386 |
| 14 | WTD | 1.761 |
| 15 | CTI | 1.249 |
| 16 | TPI | 0.936 |
| 17 | WBE | 0.713 |

## Binary Classification Modeling

### Data-Driven Classification Thresholds

Based on the basic descriptive statistics of the selected features, presented in Table 7, a few data-driven thresholds were chosen for post-classification optimization of the results. The selected thresholds are shown in bold red in the Table 7. These data-driven thresholds, plus the Mw 5 threshold for liquefaction triggering (10 variables in total), results in pre-defined non-liquefaction labels for 9418 samples out of 247114 total samples (3.8 % of the samples).

**Table 7.** Descriptive data statistics of the continuous geospatial variables by class.

| *Variable* | *Class* | *Min* | *Max* | *Mean* | *Median* | *Mode* | *STD* | *Skew* |
|---|---|---|---|---|---|---|---|---|
| ***$Vs_{30}$*** | Liq. | 137.22 | 900.00 | 232.29 | 219.32 | 260.20 | 78.85 | 2.93 |
| | Non-Liq. | 98.00 | 900.00 | 460.36 | 434.02 | 641.30 | 193.00 | 0.48 |
| ***Elevation*** | Liq. | 0.00 | 3649.99 | 11.96 | 3.82 | 0.00 | 37.40 | 26.11 |
| | Non-Liq. | 0.00 | 4130.48 | 346.85 | 184.57 | 0.00 | 413.91 | 1.99 |
| ***Thickness*** | Liq. | 0.00 | 50.00 | 44.71 | 50.00 | 50.00 | 11.49 | -2.37 |
| | Non-Liq. | 0.00 | 50.00 | 13.52 | 3.25 | 1.00 | 18.06 | 1.21 |
| ***TPI*** | Liq. | **-49.03** | **44.57** | -0.19 | -0.08 | 0.00 | 1.73 | -3.95 |
| | Non-Liq. | -91.50 | 118.75 | 0.07 | -0.12 | 0.00 | 11.73 | 0.40 |
| ***TRI*** | Liq. | 0.00 | **322.23** | 6.57 | 4.07 | 0.00 | 9.82 | 7.46 |
| | Non-Liq. | 0.00 | 596.15 | 64.17 | 40.44 | 0.00 | 69.03 | 1.60 |
| ***CTI*** | Liq. | 505.46 | 2731.87 | 1018.14 | 961.09 | 928.39 | 219.13 | 1.73 |
| | Non-Liq. | 386.38 | 2471.00 | 829.72 | 797.58 | 857.00 | 211.61 | 1.16 |
| ***PGV*** | Liq. | **1.01** | 158.15 | 45.46 | 33.79 | 28.54 | 28.31 | 0.96 |
| | Non-Liq. | 0.04 | 166.81 | 21.64 | 17.26 | 20.73 | 17.97 | 2.02 |
| ***Magnitude*** | Liq. | 5.80 | 9.10 | 7.66 | 7.70 | 7.70 | 0.96 | 0.26 |
| | Non-Liq. | 4.00 | 9.10 | 7.40 | 7.60 | 7.90 | 1.02 | -0.53 |
| ***DC*** | Liq. | 0.00 | 356.15 | 27.00 | 12.19 | 0.00 | 29.35 | 1.18 |
| | Non-Liq. | 0.00 | 370.99 | 58.69 | 28.12 | 78.00 | 76.16 | 1.95 |
| ***DR*** | Liq. | 0.00 | 1287.21 | 25.32 | 15.18 | 0.00 | 26.30 | 7.04 |
| | Non-Liq. | 0.00 | 1301.71 | 76.89 | 32.12 | 0.00 | 134.18 | 5.54 |
| ***DL*** | Liq. | 0.00 | 133.07 | 10.57 | 8.70 | 0.00 | 8.45 | 1.19 |
| | Non-Liq. | 0.00 | 142.27 | 18.11 | 13.53 | 0.00 | 16.53 | 2.51 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ***DWB*** | Liq. | 0.00 | **40.39** | 4.60 | 2.56 | 0.00 | 4.62 | 1.27 |
| | Non-Liq. | 0.00 | 109.33 | 10.03 | 7.41 | 0.00 | 11.37 | 4.35 |
| ***WBE*** | Liq. | 0.00 | **1651.00** | 8.38 | 0.80 | 0.00 | 45.01 | 17.49 |
| | Non-Liq. | 0.00 | 2703.00 | 196.00 | 25.00 | 0.00 | 317.63 | 2.52 |
| ***HWB*** | Liq. | **-1476.11** | **2342.99** | 3.73 | 1.93 | 0.00 | 16.98 | 25.62 |
| | Non-Liq. | -2800.10 | 3212.48 | 105.10 | 51.68 | 0.00 | 340.80 | -0.87 |
| ***HAND*** | Liq. | 0.00 | **1810.75** | 5.85 | 4.00 | 4.00 | 10.75 | 61.37 |
| | Non-Liq. | 0.00 | 4138.25 | 108.08 | 43.75 | 2.00 | 161.71 | 3.10 |
| ***WTD*** | Liq. | 0.00 | **216.13** | 1.69 | 0.63 | 0.00 | 4.47 | 9.23 |
| | Non-Liq. | 0.00 | 488.48 | 31.35 | 20.19 | 0.00 | 36.89 | 2.35 |
| ***AI*** | Liq. | 0.00 | **27659.90** | 7978.74 | 6159.70 | 11242.63 | 5595.91 | 0.47 |
| | Non-Liq. | 0.00 | 39830.00 | 10264.86 | 10444.00 | 0.00 | 6376.04 | 0.42 |

**Voting Classification System**

The voting classification system was developed by training individual classifiers and aggregating them through hard and soft voting. The individual earthquake events were tested via the leave-one-out approach, and the accuracy results including, overall accuracy, sensitivity, specificity, and area under the ROC curve (AUC) are provided in Table 8 for all earthquakes in the inventory, ranked by the validation's AUC value, except for the non-liquefaction events which come at the bottom of the table, ranked by their overall accuracy. The prevalence column in Table 6 shows the percentage of liquefaction points in the inventory for each event. The accuracy of predictions for the 2003 San Simeon and the 2010 Haiti earthquakes were the best among the earthquakes with liquefaction observations, followed by the 2010 Tottori. Figure 6 shows the predicted liquefaction maps for these 3 events.

**Table 8.** Accuracy statistics of the leave-one-out approach. All values are in percentage (%).

| Rank | Earthquake | Prevalence | Overall Accuracy | Specificity | Sensitivity | AUC |
|---|---|---|---|---|---|---|
| 1 | **San Simeon** | 7.79 | 96.31 | 99.11 | 63.16 | 98.77 |
| 2 | **Haiti** | 2.00 | 83.71 | 83.38 | 100.00 | 97.36 |
| 3 | **Tottori** | 49.43 | 91.96 | 84.63 | 99.45 | 96.65 |
| 4 | **Christchurch** | 95.59 | 93.22 | 83.15 | 93.69 | 95.99 |
| 5 | **Illapel** | 0.77 | 98.93 | 99.23 | 60.00 | 95.64 |
| 6 | **Samara** | 1.48 | 86.64 | 86.56 | 92.31 | 95.49 |
| 7 | **Loma Prieta** | 28.16 | 82.32 | 75.79 | 98.99 | 94.82 |
| 8 | **Cephalonia** | 5.31 | 96.46 | 99.53 | 41.67 | 94.68 |
| 9 | **Nihonkai** | 79.22 | 92.40 | 80.90 | 95.41 | 93.86 |
| 10 | **Kobe** | 76.11 | 91.36 | 67.70 | 98.78 | 92.40 |
| 11 | **Kocaeli** | 1.96 | 95.24 | 96.17 | 48.57 | 92.17 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 12 | **Honduras** | 2.40 | 54.90 | 53.79 | 100.00 | 92.14 |
| 13 | **Nisqually** | 8.45 | 64.40 | 61.78 | 92.86 | 91.67 |
| 14 | **Hokkaido** | 29.42 | 89.41 | 93.13 | 80.48 | 90.96 |
| 15 | **Nepal (Gorkha)** | 1.91 | 93.80 | 94.81 | 41.67 | 90.50 |
| 16 | **Muisne** | 1.10 | 85.40 | 85.32 | 92.31 | 89.65 |
| 17 | **Nigata 1964** | 74.61 | 88.62 | 68.80 | 95.36 | 89.60 |
| 18 | **Miyagi ken** | 5.93 | 65.56 | 63.52 | 98.08 | 89.42 |
| 19 | **Darfield** | 95.45 | 91.13 | 67.03 | 92.28 | 88.92 |
| 20 | **Tokachi** | 3.23 | 69.35 | 68.62 | 91.16 | 88.88 |
| 21 | **Napa** | 0.72 | 56.16 | 55.84 | 100.00 | 88.69 |
| 22 | **Tohoku** | 89.19 | 90.72 | 54.91 | 95.06 | 86.86 |
| 23 | **Oklahoma** | 0.74 | 92.96 | 93.25 | 50.00 | 86.47 |
| 24 | **Kumamoto** | 26.09 | 81.18 | 84.32 | 72.31 | 85.89 |
| 25 | **Nigata 2004** | 80.00 | 77.66 | 85.47 | 75.70 | 85.35 |
| 26 | **Chiba** | 47.37 | 72.47 | 54.91 | 91.99 | 84.84 |
| 27 | **Duzce** | 0.48 | 92.82 | 92.95 | 66.67 | 84.80 |
| 28 | **Nigata 2007** | 27.81 | 66.47 | 53.55 | 100.00 | 83.19 |
| 29 | **Tecoman** | 1.05 | 69.44 | 69.35 | 77.78 | 81.84 |
| 30 | **Puget Sound 1949** | 1.92 | 92.31 | 93.03 | 55.56 | 81.62 |
| 31 | **Telire Limon** | 3.00 | 59.53 | 58.35 | 97.67 | 81.48 |
| 32 | **Pisco** | 2.50 | 97.84 | 99.83 | 20.00 | 80.06 |
| 33 | **Denali** | 0.83 | 94.34 | 95.05 | 10.00 | 75.29 |
| 34 | **Achaia** | 1.06 | 85.89 | 86.35 | 42.86 | 73.38 |
| 35 | **Meinong** | 1.99 | 39.17 | 37.94 | 100.00 | 73.36 |
| 36 | **Emilia** | 6.74 | 51.86 | 48.88 | 93.10 | 72.99 |
| 37 | **Baja California** | 2.52 | 51.66 | 50.62 | 92.00 | 67.56 |
| 38 | **Puget Sound 1965** | 3.73 | 84.14 | 85.97 | 37.12 | 65.18 |
| 39 | **Northridge** | 2.33 | 97.67 | 99.72 | 11.90 | 62.42 |
| 40 | **Arequipa** | 0.74 | 98.52 | 99.25 | 0.00 | 62.18 |
| 41 | **Maule** | 0.53 | 94.21 | 94.62 | 16.67 | 58.58 |
| 42 | **Chi-Chi** | 2.56 | 70.27 | 94.23 | 0.48 | 50.00 |
| 43 | **Aquila** | 1.23 | 98.77 | 100.00 | 0.00 | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| 44 | **Iquique** | 1.25 | 98.75 | 100.00 | 0.00 | - |
| 45 | **Wenchuan** | 1.38 | 98.48 | 99.86 | 0.00 | - |
| 46 | **VanTab** | 2.84 | 97.16 | 100.00 | 0.00 | - |
| 47 | **Virginia** | 0.57 | 99.43 | 100.00 | 0.00 | - |
| 1 | **Chino Hills** | 0.00 | 100.00 | - | - | - |
| 1 | **Hector Mine** | 0.00 | 100.00 | - | - | - |
| 1 | **Piedmont** | 0.00 | 100.00 | - | - | - |
| 1 | **Yountville** | 0.00 | 100.00 | - | - | - |
| 5 | **Central Italy** | 0.00 | 99.78 | - | - | - |
| 6 | **Iwate** | 0.00 | 89.43 | - | - | - |

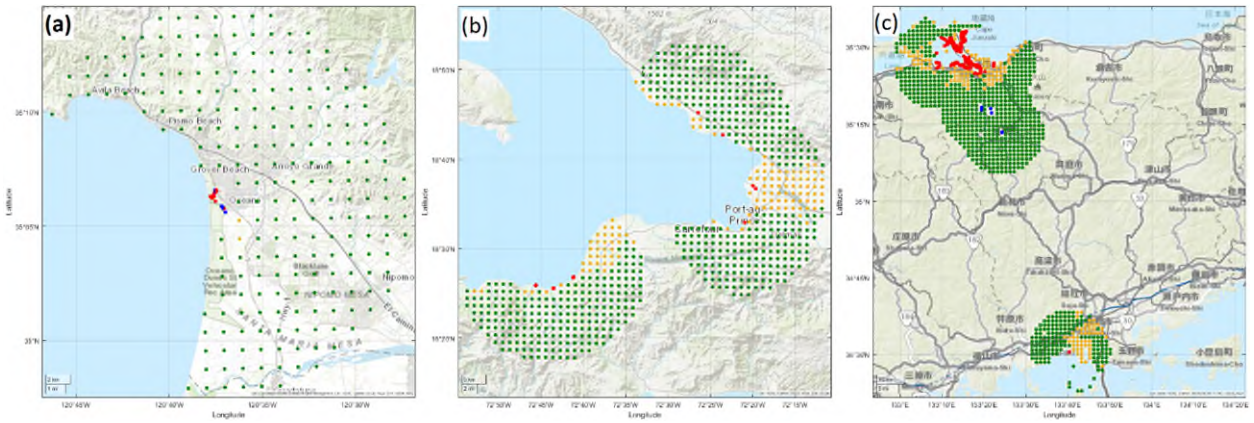


**Figure 6.** a) Predicted liquefaction map of the 2003 San Simeon; b) 2010 Haiti; and c) 2010 Tottori earthquakes. Red is correctly predicted liquefaction. Green is correctly predicted non-liquefaction. Yellow is wrongly predicted liquefaction. Blue is wrongly predicted non-liquefaction.

The 2011 Christchurch earthquake has the fourth best accuracy among the liquefaction-observed earthquakes in the inventory, while the 2010 Darfield event's accuracy is relatively lower (ranked 19$^{th}$ among the liquefaction-observed events). Figure 7 provides the liquefaction maps for these two events (Canterbury 2010-11 events in New Zealand). The difference in the predicted maps' accuracy was also observed in the results of Todorovic and Silva (2022) for their random forest (decision tree-based) model, but with a more significant difference. An interesting observation here is that the decision tree classifier (coastal classifier 1) in this study also under-performed compared to the other 5 classifiers in the voting classification system. The voting system has a higher accuracy (add accuracy) than all individual 6 classifiers, which demonstrates the integrated power of the proposed method in providing more accurate predictions. Table 9 shows the accuracy (AUC) of individual classifiers, in addition to the final voting system for all tested earthquakes.

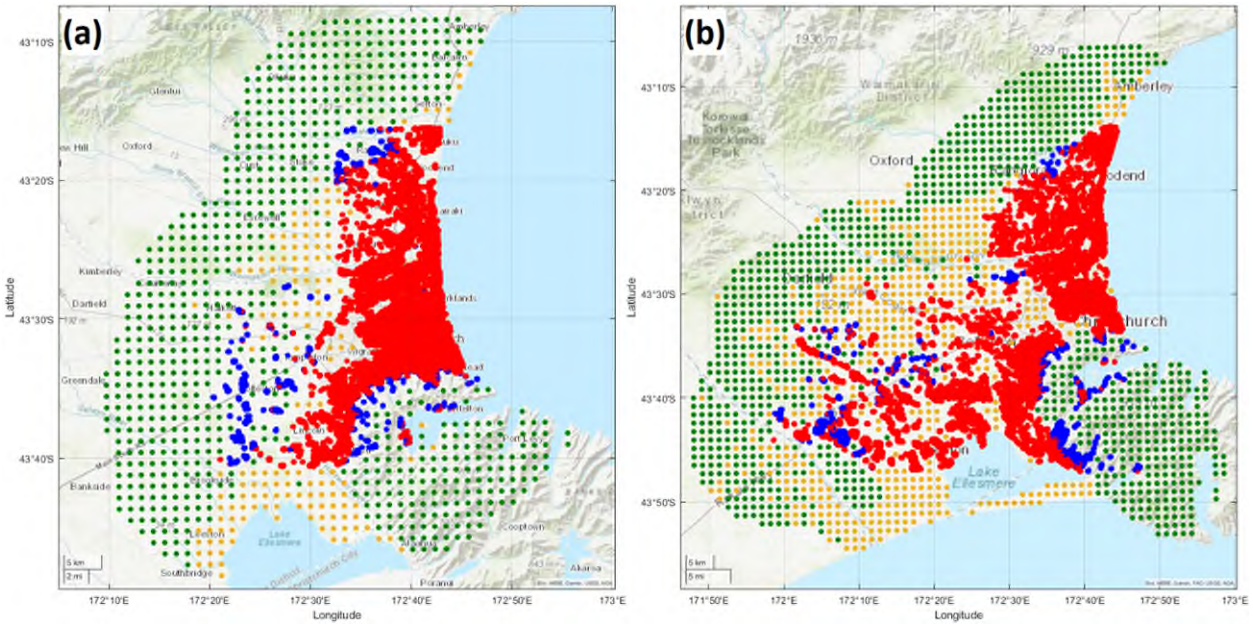


**Figure 7.** a) Predicted hazard map of the 2011 Christchurch; and b) 2010 Darfield earthquakes. Red is correctly predicted liquefaction. Green is correctly predicted non-liquefaction. Yellow is wrongly predicted liquefaction. Blue is wrongly predicted non-liquefaction.

According to Table 9, the AUC value for the proposed voting system is higher than all individual classifiers in 88.4 % of the tested earthquakes for which AUC is calculated. For the majority of other events, it's very close to the highest AUC among the classifiers. This is evidence of the benefit of the proposed method, especially compared to the three global classifiers, which represent statistically optimal global models. Additionally, the spatial extent of liquefaction is moderated, even in the cases where the voting system shows a relatively lower AUC than the highest performing individual classifier. A very important observation is that not a single global classifier can be considered as best-performing for all the events. According to the results provided in Table 9, among the 3 global models, global classifiers 1, 2, and 3, perform the best on 30 %, 35 % and 30 % of the events, respectively, while they perform equally poor on the remaining 5 % of the events. The voting classification system performs better by taking into account the decisions made by coastal/non-coastal and regional classifiers in addition to the global models.

**Table 9.** Comparative accuracy of single classifiers versus the proposed system of voting classifiers. All values are AUC in percentage (%). The events are listed in alphabetical order.

| Earthquake | Global 1 | Global 2 | Global 3 | Coastal / Non-Coastal 1 | Coastal / Non-Coastal 2 | Regional | Proposed Method |
|---|---|---|---|---|---|---|---|
| **Achaia** | 66.37 | 75.51 | 68.44 | 64.76 | 51.93 | 60.23 | 73.38 |
| **Arequipa** | 50.00 | 50.00 | 50.00 | 53.94 | 50.00 | 61.82 | 62.18 |
| **Baja California** | 50.00 | 68.93 | 50.00 | 50.00 | 59.19 | 69.34 | 67.56 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Cephalonia** | 62.27 | 70.83 | 54.17 | 89.80 | 62.50 | 86.33 | 94.68 |
| **Chi-Chi** | 50.00 | 50.00 | 50.00 | 59.18 | 75.77 | 50.00 | 73.95 |
| **Chiba** | 72.86 | 65.23 | 77.82 | 70.06 | 58.82 | 69.15 | 84.84 |
| **Christchurch** | 88.57 | 83.26 | 82.76 | 62.77 | 69.90 | 89.84 | 95.99 |
| **Darfield** | 78.80 | 73.72 | 75.66 | 57.40 | 71.54 | 72.97 | 88.92 |
| **Denali** | 50.00 | 50.00 | 52.67 | 72.35 | 67.11 | 51.46 | 75.29 |
| **Duzce** | 50.00 | 70.03 | 50.00 | 50.00 | 75.80 | 80.05 | 84.80 |
| **Emilia** | 55.08 | 58.36 | 72.75 | 49.24 | 55.94 | 61.60 | 72.99 |
| **Haiti** | 89.99 | 90.26 | 89.51 | 89.72 | 93.12 | 97.82 | 97.36 |
| **Hokkaido** | 89.20 | 85.84 | 72.86 | 67.68 | 85.95 | 77.21 | 90.96 |
| **Honduras** | 78.88 | 78.98 | 86.74 | 76.52 | 76.70 | 78.50 | 92.14 |
| **Illapel** | 50.00 | 88.92 | 50.00 | 79.30 | 50.00 | 82.53 | 95.64 |
| **Kobe** | 81.80 | 82.68 | 88.06 | 86.86 | 80.62 | 88.59 | 92.40 |
| **Kocaeli** | 68.40 | 84.20 | 54.20 | 77.71 | 50.00 | 85.82 | 92.17 |
| **Kumamoto** | 83.73 | 81.77 | 66.06 | 74.63 | 76.02 | 79.89 | 85.89 |
| **Loma Prieta** | 86.59 | 87.67 | 93.13 | 50.00 | 85.53 | 87.89 | 94.82 |
| **Maule** | 58.12 | 57.18 | 54.13 | 55.09 | 54.54 | 50.00 | 58.58 |
| **Meinong** | 69.06 | 69.15 | 67.03 | 66.38 | 66.67 | 62.25 | 73.36 |
| **Miyagi ken** | 79.25 | 80.10 | 84.11 | 82.65 | 76.38 | 82.89 | 89.42 |
| **Muisne** | 70.27 | 82.79 | 67.74 | 89.11 | 87.71 | 53.92 | 89.65 |
| **Napa** | 78.10 | 79.74 | 85.58 | 79.93 | 74.45 | 79.93 | 88.69 |
| **Nepal (Gorkha)** | 84.60 | 59.74 | 50.00 | 78.88 | 69.29 | 50.00 | 90.50 |
| **Nigata 1964** | 81.85 | 82.29 | 81.82 | 82.50 | 80.53 | 83.46 | 89.60 |
| **Nigata 2004** | 82.28 | 80.64 | 56.77 | 62.49 | 80.43 | 80.87 | 85.35 |
| **Nigata 2007** | 76.09 | 75.20 | 78.31 | 76.19 | 76.52 | 77.39 | 83.19 |
| **Nihonkai** | 87.70 | 89.32 | 70.31 | 85.05 | 87.19 | 88.25 | 93.86 |
| **Nisqually** | 75.93 | 77.02 | 87.09 | 88.04 | 74.16 | 63.34 | 91.67 |
| **Northridge** | 50.00 | 62.31 | 50.00 | 54.79 | 50.00 | 56.80 | 62.42 |
| **Oklahoma** | 71.55 | 59.79 | 50.00 | 58.58 | 93.84 | 50.00 | 86.47 |
| **Pisco** | 50.00 | 79.32 | 50.00 | 74.53 | 50.00 | 59.15 | 80.06 |
| **Puget Sound 1949** | 77.85 | 74.23 | 66.27 | 62.51 | 74.82 | 73.93 | 81.62 |
| **Puget Sound 1965** | 61.61 | 60.31 | 59.90 | 59.82 | 58.88 | 64.29 | 65.18 |
| **Samara** | 88.74 | 82.32 | 92.85 | 88.97 | 89.14 | 92.10 | 95.49 |
| **San Simeon** | 50.00 | 95.11 | 55.26 | 71.24 | 73.46 | 97.33 | 98.77 |
| **Tecoman** | 74.57 | 80.60 | 74.22 | 63.69 | 72.79 | 76.57 | 81.84 |
| **Telire Limon** | 77.69 | 76.95 | 64.97 | 77.31 | 77.94 | 60.00 | 81.48 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Tohoku** | 72.96 | 67.93 | 70.18 | 69.59 | 70.21 | 76.21 | 86.86 |
| **Tokachi** | 78.15 | 77.21 | 85.65 | 83.21 | 78.12 | 83.98 | 88.88 |
| **Tottori** | 90.67 | 90.87 | 94.87 | 91.99 | 91.83 | 94.12 | 96.65 |
| **Wenchuan** | 50.00 | 50.00 | 50.00 | 52.60 | 63.79 | 50.00 | 65.41 |

## Section 3. Pixel-based Classification Method for Earthquake-Induced Landslide Mapping Using Imagery, Geospatial Data and Temporal Change Information

(The following content has been excerpted from a peer reviewed publication with the same title and the following authors)

Adel Asadi[1], Laurie G. Baise[1], Magaly Koch[2], Babak Moaveni[1], Snehamoy Chatterjee[3], Yusupujiang Aimaiti[2]

[1] Geohazards Research Lab, Department of Civil and Environmental Engineering, School of Engineering, Tufts University, Medford, MA 02155, USA

[2] Center for Remote Sensing, Department of Earth and Environment, Boston University, Boston, MA 02215, USA

[3] Geological and Mining Engineering and Sciences Department, Michigan Technological University, Houghton, MI 49931, USA

### Abstract

A series of earthquakes hit Kumamoto, Japan, over a period of two days in April 2016, which caused numerous landslides. In this study, high-resolution pre-event and post-event optical imagery, plus bi-temporal Synthetic Aperture Radar (SAR) data are paired with geospatial data to train a pixel-based machine learning classification algorithm using logistic regression to identify landslides occurred because of the Kumamoto earthquakes. The geospatial data used include a categorical variable (surficial geology), and continuous variables including elevation, slope, aspect, curvature, annual precipitation, and landslide probability derived by the USGS preferred geospatial model which incorporates ground shaking in the input parameters. Grayscale index change and vegetation index change are also calculated from the optical imagery and used as input variables, in addition to temporal differences in HH and HV amplitudes of SAR data. A detailed human-drawn landslide occurrence inventory was used as ground-truth for model development and testing. The selection of optimal features was done using a supervised feature ranking method based on the Receiver Operating Characteristic (ROC) curve. To weigh the benefit of combining different types of imagery, temporal change information and geospatial environmental indicators for landslide mapping after earthquakes, five different combinations of features were tested, and the results showed that adding data of selected geospatial parameters (landslide probability, slope, curvature, precipitation, and geology) plus selected change indices (grayscale index change, vegetation index change, and HV amplitude difference of SAR data) to the imagery (post event optical) lead to the highest classification accuracy of 86.5% on class-balanced independent testing data.

## Methodology

Given the available data from the Kumamoto Earthquake 2016, a methodology for feature selection followed by a pixel-based binary classification method is presented in this study. The feature selection step helps to determine which data inputs help the model to distinguish better between the two classes (landslide and non-landslide) without overfitting, since one way to reduce the overfitting is to reduce the dimensionality of the data. Feature analysis can, therefore, help us understand variable importance in the classification modeling.

In this study, landslide mapping is a binary classification problem, which means that the goal of developing a model is to predict each pixel as either positive (landslide) or negative (non-landslide) for landslide mapping. The flowchart of the proposed methodology is shown in Figure 8. In this section, data processing and modeling is presented via subsections starting with the data normalization as a pre-processing step and ending with model accuracy evaluation and post-processing steps.

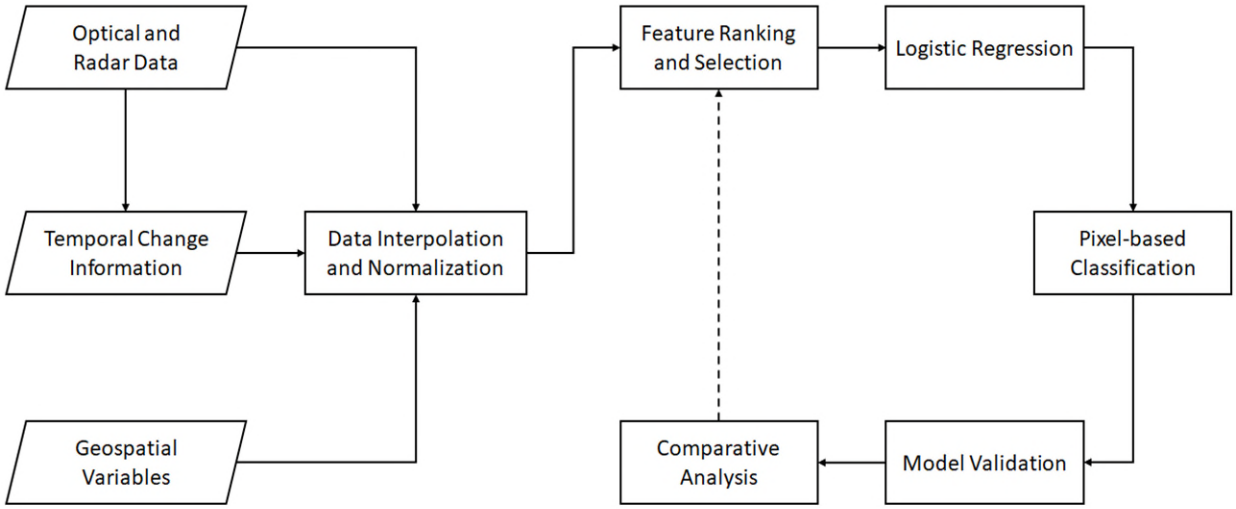


Figure 8. Flowchart of the proposed landslide mapping algorithm.

## Data Normalization

After the resampling and stacking of the datasets, all the data layers were normalized as a pre-processing step. Since the geology variable is originally categorical, but categorical data is not desirable for the classification method used in this research, the geological categories are sorted by the percentage of landslide occurrence in the categories within the training data, and the percentage of landslide pixels is substituted with the category numbers. This leads to a meaningful set of numbers for the geology variable, where the higher number indicates a higher probability of landslide occurrences. Then, the geology feature is considered same as other continuous variables, and the variable is normalized as well.

Both the feature selection and classification models are developed based on the normalized data of the explanatory variables or data inputs, $\boldsymbol{X}$, which significantly reduces the impact of data range on the final landslide prediction model and increases the processing speed with fast convergence. Using the mean ($\mu$) and standard deviation ($\sigma$) of individual explanatory variables,

data of all variables were normalized using Equation 11 to have a mean of 0 and standard deviation of 1.

$$z_i = \frac{x_i - \mu}{\sigma} \tag{11}$$

**Feature Ranking**

Different combinations of features are used in the model validation step based on the feature ranking results, to perform a comparative analysis and to choose the best performing model based on the accuracy of the results for each combination.

The implemented algorithm ranks the features in data using an independent evaluation criterion for binary classification. Assume that $Z$ is a matrix where every row is an observed vector ($Z_i$), and the number of columns correspond to the original number of features. The criterion for feature ranking is the area under the empirical receiver operating characteristic (ROC) or the Area Under the Curve (AUC) (Theodoridis 1999). The receiver operating characteristic (ROC) curve is a graphical plot that illustrates the diagnostic ability of the binary classification system. It is generated by plotting the true positive rate for a given classifier against the false positive rate for various thresholds. For a perfect classifier, AUC = 1. For a classifier that randomly assigns observations to classes, AUC = 0.5 (Fawcett 2004).

ROC curves and AUC scores also allow us to compare the performance of different classifiers for the same problem. The ROC curve is produced by calculating and plotting the true positive rate against the false positive rate for a single classifier at a variety of thresholds between 0 and 1. The optimal threshold leads to the highest AUC and highest classification accuracy. At the feature ranking step in this research, each variable $Z_i$ is tested through various thresholds within its range (min-max) to calculate TPR and FPR for those thresholds, which are needed to generate the ROC curve for the feature.

**Binary Classification Modeling**

In this subsection, different aspects of the classification modeling are explained in detail.

**Logistic Regression**

The pixel-based binary classification step uses logistic regression algorithm to predict landslide and non-landslide pixels. Logistic Regression is a supervised machine learning algorithms which is commonly used for binary class classification problems. Logistic regression was borrowed by machine learning from the field of statistical models and is currently one of the most well-known and widely used machine learning algorithms (James et al. 2013). Logistic regression operates using a sigmoidal function for values ranging between 0 and 1. In the context of image processing, this could mean identifying whether a given pixel belongs to a particular class.

A logistic regression algorithm takes as its input a feature vector $Z_i$ and outputs a probability, $\hat{y}$=P(y=1|z), where the feature vector represents an object belonging to the class *y*. For image processing, the feature vector might be just the values of the red, green, and blue (RGB) channels for each pixel in the image, a one-dimensional array of real numbers formed by flattening the three-dimensional array of RGB pixel values. A logistic regression model is described by the logistic (sigmoid) function shown via Equation 12.

$$\sigma(w) = \frac{1}{1+e^{-w}} \quad (12)$$

and $w$ contains the weights (coefficients) and the bias (intercept) of the linear equation. To calculate the coefficients in a way that the error is minimized, a loss error function is defined as shown via Equation 13, which describes how far the estimates are from actual observations.

$$L(\hat{y}, y) = \frac{(\hat{y}-y)^2}{2} \quad (13)$$

**Data Partitioning**

To train the classifier, it is recommended to separate data samples into distinct groups for model development and model testing, to avoid overfitting or underfitting. The difference in accuracies between the training and test sets can indicate whether a model is overfitting. The model is overfitting when the difference is large and the accuracy on the training set is very high. Underfitting, on the other hand, occurs when a model is not sophisticated enough to precisely capture relationships between predictors and responses, resulting in low accuracy on both training and test sets.

In this study, a special sampling process is designed to create model development and testing data. The area of study was split into 10-by-20 grid regions (100 model development and 100 model testing regions). The resulting dataset was significantly unbalanced with only 3.32% of pixels belonging to the landslide class. Class imbalance can cause overfitting in the classification algorithm as the model is rewarded by only predicting the majority class. To balance the datasets, only 10% of landslide points in the model development grid regions were sampled and added to only 0.4% of non-landslide pixels in those grid regions to form the sampled and balanced model development dataset.

Based on the sampling strategy, the model development data of the landslide class contains only 5.67% of the total pixels assigned as landslide pixels in the study area, which means that the final model has not ever seen 94.33% of ground-truth landslide samples. The reason for the sampling process is to highlight the fact that no significant amount of data is needed to train the algorithms. This will also be helpful in a trustworthy validation step by considering most pixels in the final predicted landslide map of the study area as new independent predictions. In addition, results of the study by Rashidian et al. (2020) on pixel-based Earthquake-induced liquefaction modeling showed that although increasing the number of training pixels increases the accuracy and reliability of the classifier, the improvement will not be meaningful after reaching a specific number of pixels (2000 pixels recommended in their study).

Figure 9 shows the ground-truth label based on the NIED landslide inventory, and the regional distinction for sampling model development and testing data, which has been implemented to avoid model bias. As shown in the Figure, model development data points have been sampled from blue boxes, whereas the testing points were sampled from the gray boxes. The testing dataset was also randomly sampled to create a balanced testing dataset consisting of all landslide pixels plus the same number of pixels from the non-landslide class (3.92% of available non-landslide pixels in the testing grid regions).

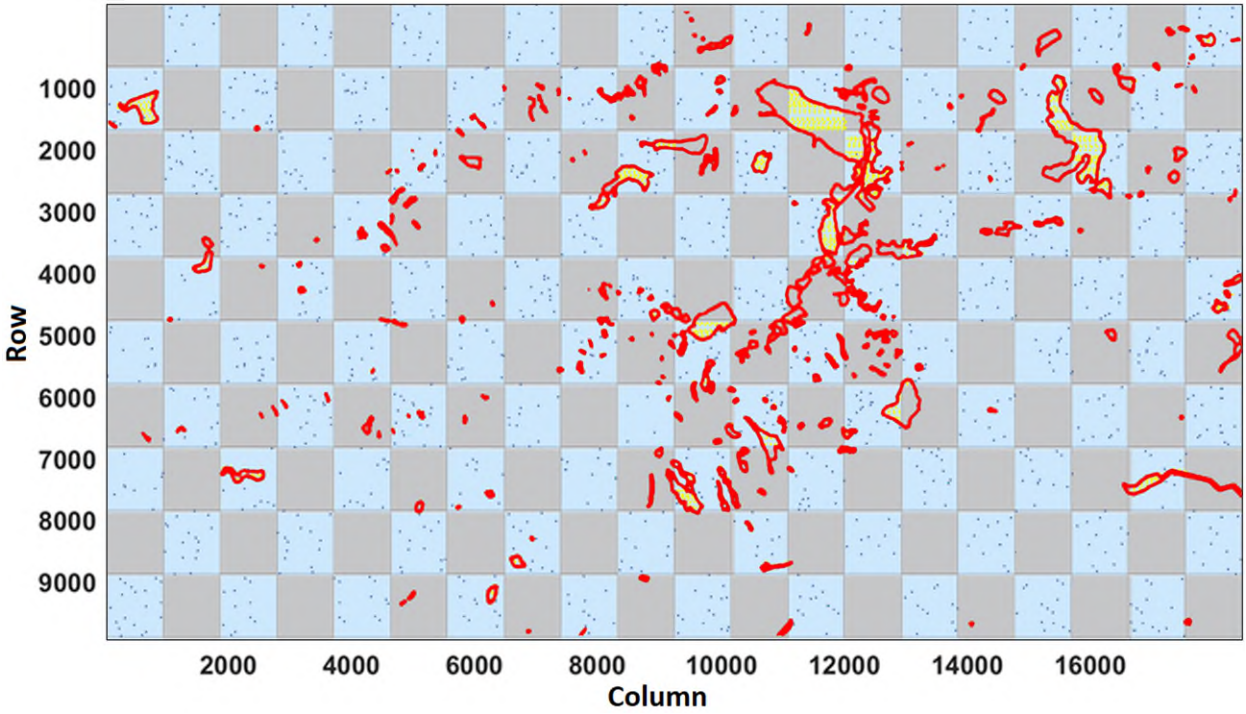


**Figure 9.** Ground-truth binary landslide label, with distinct model development and testing regions. Model development regions are shown via light blue, and testing regions are shown via gray color. Yellow dots are the sampled training data for the landslide class, and blue dots are the sampled training data for the non-landslide class. Rows and columns as axes titles are indicators of pixel number (pixel size is 0.5 m). The sample points are schematical, and are not real indicator of the sampled data, since the points are very small, and needed to be enlarged for visualization. The number of sampled data is higher that visualized points in this figure.

## Model Evaluation

After random sampling of the model development data, the collected samples are divided into two groups of training and validation. The validation process was implemented by hold-out validation of 25% of the sampled and balanced model development data to help the training process. The validation accuracy is reported for each of the evaluated case models. Then, the distinct testing data samples from testing grid regions are used to test the predictive ability of the classification model on an independent unseen dataset, and to calculate the accuracy indices explained below.

There are several classification accuracy indices which were used in this study to make sure the model performs well. The proportion of correctly classified observations per positive class is referred to as the Sensitivity or True Positive Rate (TPR). Specificity or True Negative Rate (TNR) is the rate of correct prediction in the negative class. The sensitivity and specificity formulas are provided via Equations 14 and 15. The proportion of incorrectly classified observations per true class is referred to as the False Negative Rates (FNR). The Area Under Curve (AUC) curve depicts the true positive versus false positive rate. The AUC number is a measure of the overall quality of the classifier. Higher AUC (maximum is 1) values indicate better classification performance.

$$Sensitivity = TPR = \frac{TP}{TP+FN} \tag{14}$$

$$Specificity = TNR = \frac{TN}{TN+FP} \quad (15)$$

The proportion of correctly classified observations in the positive (landslide) class is referred to as the Positive Predictive Values (PPV) or precision. A high precision indicates that the model has a high probability of correctly classifying positive samples. The recall or Negative Predictive Value (NPV) is proportion of correctly classified observations in the negative (non-landslide) class. While the recall explains how sensitive the model is to identifying actual positive samples, it also quantifies the probability of detecting actual positive samples. Because precision and recall assess different aspects of the model, an index that combines the two is also used. The F1 score is the harmonic mean of the precision and recall, where an F1 score reaches its best value at 1 and worst at 0. The precision and recall formulas are provided via Equations 16 and 17.

$$Precision = PPV = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (16)$$

$$Recall = NPV = \frac{\mathrm{TN}}{\mathrm{TN+FN}} \quad (17)$$

**Landslide Map Post-Processing**

After generating the final landslide maps, to remove the noise and smooth the maps, two-dimensional median filtering was performed on the generated maps, which substitutes isolated pixels with the dominant class around them. Five rounds of smoothing are performed to avoid pixelated landslide maps, and to smooth the boundaries of the features. The window size selected for the post-processing step is 21×21 pixels in this study as a moderate size window as a trade-off to isolate and filter noise pixels while not losing very small-size landslide features. To analyze the impact of post-processing on the landslide map in terms of accuracy statistics, the generated map is compared with the ground-truth label, and the map overall accuracies before and after the smoothing are reported.

**Feature Importance Analysis**

Feature importance is also assessed by multiplying the coefficients of the logistic regression model by the standard deviation of the corresponding continuous parameters. Since all variables were normalized in the pre-processing step, having standard deviation of 1, the coefficients of the logistic regression model themselves are also the indicators for the feature importance in this study.

**Analysis Results**

In this section, the results of feature ranking step and the binary classification modeling are provided. Model validation is performed, and the visualizations are provided plus a discussion of the methods and outputs.

**Feature Selection**

In this study, it was decided to keep all RGB color channels of the post-event imagery active in the training process for the classification models. Feature ranking of the geospatial features and the temporal change features was performed to evaluate different combinations. The goal of performing this step is to select high-weight features per the categories of change information and geospatial information. The supervised feature ranking results via ROC calculation are shown in

Table 10. The weights of features in the landslide and non-landslide classes show that the two change bands of vegetation difference and grayscale change have the highest impact on the classification as compared to the other variables. The third high-weight band in the change information category is HV amplitude difference. Geology, slope, precipitation, curvature, and landslide probability are the other high-weight geospatial features. The criteria to select the high-weight features was to have a weight of at least 0.55 in the feature ranking step (Table 10). It should be noted that having a weight of 0.5 is indicator of a random classifier.

**Table 10.** Feature ranking results by ROC scores for the continuous variables.

| Rank | Feature Description | Feature Type | Weight |
|---|---|---|---|
| 1 | Grayscale Change | Change | 0.792 |
| 2 | Vegetation Change | Change | 0.791 |
| 3 | Geology | Geospatial | 0.783 |
| 4 | Slope | Geospatial | 0.669 |
| 5 | Precipitation | Geospatial | 0.582 |
| 6 | Curvature | Geospatial | 0.580 |
| 7 | HV Amplitude Difference | Change | 0.565 |
| 8 | Landslide Probability | Geospatial | 0.562 |
| 9 | Digital Elevation | Geospatial | 0.542 |
| 10 | HH Amplitude Difference | Change | 0.529 |
| 11 | Aspect | Geospatial | 0.524 |

## Classification and Model Evaluation

To evaluate how data fusion can aid in the development of landslide inventories from post-event imagery, different subsets of features were created with increasing complexity to test the performance of the logistic regression classification algorithm. All models used the post-event RGB imagery, including Model 1 developed solely via RGB channels. Next, a model that adds the optimal temporal change indices was tested (Model 2 = RGB + Change). Next, a model that adds the optimal geospatial features was tested (Model 3 = RGB + Geospatial). Next, a model that adds both optimal change indices and optimal geospatial features was tested (Model 4 = RGB + Change + Geospatial). Then, Model 5 was tested via all available data (14 bands).

Table 11 provides the classification accuracy results of different base models. Based on the overall classification accuracy of the model applied on the testing data, as shown in the table, the recommended model is Model 4 which uses RGB, selected change indices (grayscale change, vegetation change, and HV amplitude difference), plus selected geospatial features (geology, slope, precipitation, curvature, and landslide probability). The proposed Model 4 outperforms other models for most of the accuracy indices, especially the overall testing accuracy (86.5%). The color imagery alone (Model 1 = RGB), proved to have the worst performance among the models,

relatively. Although, it is observed that all models achieved a high level of learning and predictive performance, with some variations related to the different inputs fed to the algorithm.

**Table 11.** Classification performance results in different case models. (* is indicator of selected high-ranked feature)

| Case Model | | | Overall Classification Accuracy % | | Area Under Curve (AUC) | | Non-Landslide Testing Accuracy % | | Landslide Testing Accuracy % | |
|---|---|---|---|---|---|---|---|---|---|---|
| No. | Variable Categories | Channel Number | Validation | Testing | Validation | Testing | TNR | NPV | TPR | PPV |
| 1 | RGB | 3 | 85.00 | 82.48 | 0.92 | 0.89 | 82.85 | 82.04 | 82.11 | 82.92 |
| 2 | RGB Change* | 6 | 86.39 | 83.90 | 0.93 | 0.91 | 85.08 | 82.33 | 82.79 | 85.47 |
| 3 | RGB Geospatial* | 8 | 87.95 | 86.13 | 0.94 | 0.93 | 87.06 | 84.99 | 85.25 | 87.29 |
| 4 | RGB Change* Geospatial* | 11 | 88.63 | 86.50 | 0.95 | 0.93 | 87.99 | 84.63 | 85.11 | 88.38 |
| 5 | All Channels | 14 | 88.64 | 86.44 | 0.95 | 0.93 | 87.91 | 84.59 | 85.06 | 88.30 |

Model 4 (the preferred Model), as well as Models 3 and 5, have higher prediction accuracy, compared to Models 1 and 2, which can be the result of using geospatial parameters, especially the geology variable, which showed very high weights in the feature ranking step. Among the Models 3, 4 and 5, the Model 3 has the lowest accuracy, probably because of not exploiting the change information. Among Models 4 and 5, the preferred Model 4 has the superior performance, with slightly better accuracy results, compared to Model 5, which can be attributed to optimal feature selection in Model 4, compared to redundant features in Model 5, leading to overfitting. The model parameters show small estimation uncertainty based on the standard errors (SE), and all selected variables are statistically significant based on the t-statistic for a two-sided test with the null hypothesis that the coefficient is zero, and the near-zero associated p-values of the hypothesis test.

Figure 10 shows the outputs of Models 1 (Figure 10-a) and 4 (Figure 10-b) to visualize the impact of adding change and geospatial information to the imagery for classification. Additionally, the final output of the Model 4 based on the 0.70 probability threshold is provided in the bottom of the Figure (10-c). The modeled maps include colors red (true positive: correctly mapped landslides), yellow (false positive: wrongly mapped landslides), and blue (false negative: missing landslide features, predicted as non-landslide). True negative pixels (correctly predicted non-landslide) are not visualized.

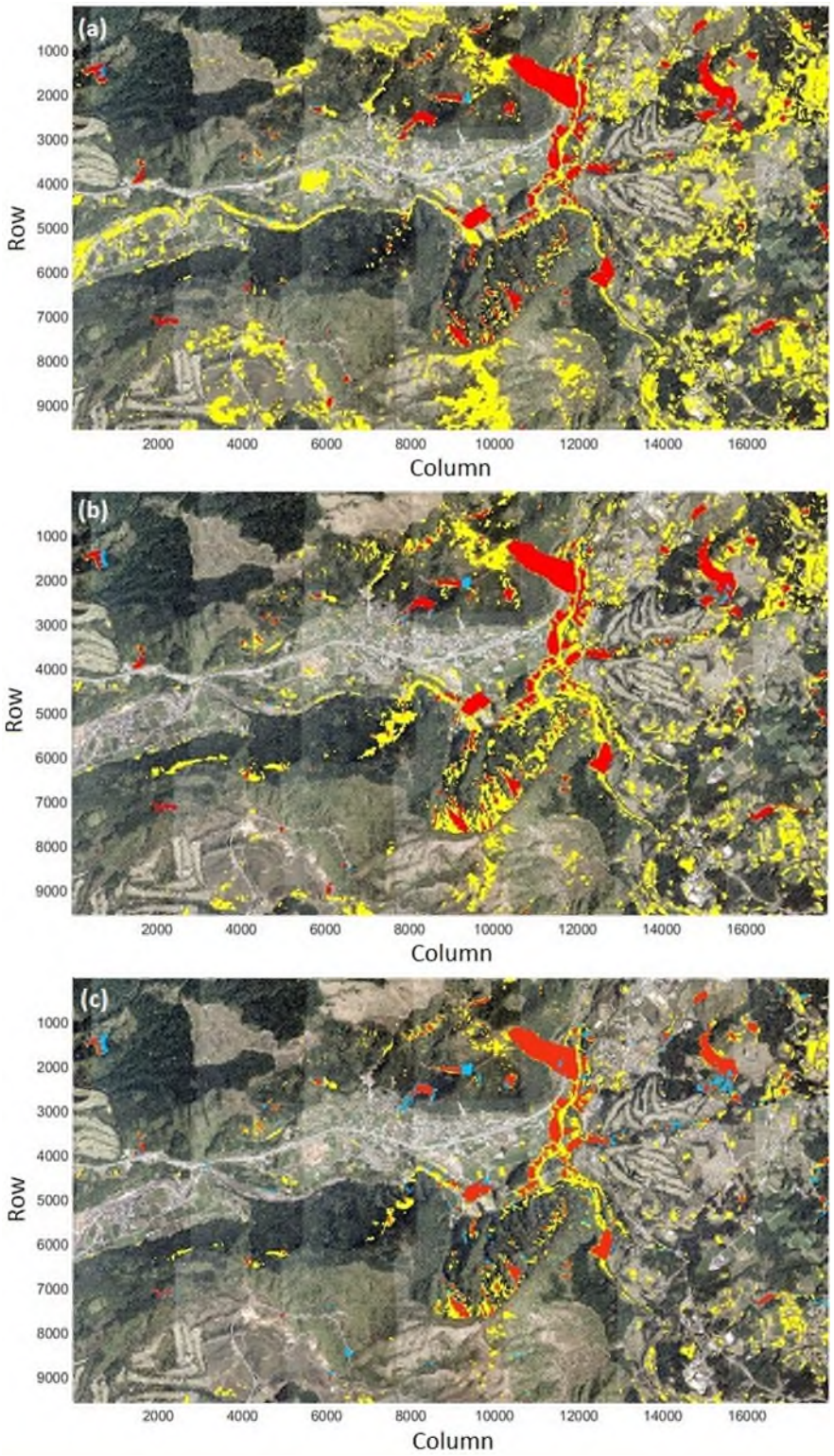


**Figure 01.** Comparative maps of the landslides by Models 1 (a), Model 4 (b) and the Model 4 with 70% probability threshold (c). The maps show post-event imagery with predicted labels. The modeled maps include colors red (true positive: correctly mapped landslides), yellow (false positive: wrongly mapped landslides), and blue (false negative: missing landslide features, predicted as non-landslide). Rows and columns as axes titles are indicators of pixel number (pixel size is 0.5 m).

## References


Allstadt, K. E., Thompson, E. M., Jibson, R. W., Wald, D. J., Hearne, M., Hunter, E. J., ... & Haynie, K. L., 2022. The US Geological Survey ground failure product: Near-real-time estimates of earthquake-triggered landslides and liquefaction. Earthquake Spectra, 38(1), 5-36. https://doi.org/10.1177/87552930211032685.

Bardet, J.P, F. Liu (2009) Motions of gently sloping ground during earthquakes. J. Geophys. Res. Earth Surf., 114 (2009), DOI:10.1029/2008JF001107.

Baise, L.G., Akhlaghi, A., Chansky, A., Meyer, M., Moaveni, B., (2021), Updating the Geospatial Liquefaction Database and Model, Final Technical Report, USGS Award #G20AP00029.

Beguería, 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.

Bozzoni, F.; Bonì, R.; Conca, D.; Lai, C.G.; Zuccolo, E.; Meisina, C., 2020, Megazonation of earthquake-induced soil liquefaction hazard in continental Europe. Bull. Earthq. Eng. 2020.

Bozzoni, F.; Bonì, R.; Conca, D.; Meisina, C.; Lai, C.G.; Zuccolo, E. A Geospatial Approach for Mapping the Earthquake-Induced Liquefaction Risk at the European Scale. Geosciences 2021, 11, 32. https://doi.org/10.3390/geosciences11010032.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. Classification and Regression Trees. Boca Raton, FL: Chapman & Hall, 1984.

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP, 2002, SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res., 16:321–357.

Christianini, N., and J. Shawe-Taylor. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge, UK: Cambridge University Press, 2000.

Cover T.; P. Hart, "Nearest neighbor pattern classification," in IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21-27, January 1967, https://doi.org/10.1109/TIT.1967.1053964.

Fernandez ´ A, García S, Galar M, Prati RC. Learning from imbalanced data sets. Springer Nature Switzerland AG; 2019. https://doi.org/10.1007/978-3-319-98074-4.

Finney, M. A., McHugh, C. W., Grenfell, I. C., Riley, K. L., & Short, K. C., 2011. A simulation of probabilistic wildfire risk components for the continental United States. Stochastic Environmental Research and Risk Assessment, 25(7), 973-1000. https://doi.org/10.1007/s00477-011-0462-z.

Fisher, R. A. "The Use of Multiple Measurements in Taxonomic Problems." Annals of Eugenics, Vol. 7, pp. 179–188, 1936. Available at https://digital.library.adelaide.edu.au/dspace/handle/2440/15227.

Furumura, T., Imai, K., & Maeda, T., 2011. A revised tsunami source model for the 1707 Hoei earthquake and simulation of tsunami inundation of Ryujin Lake, Kyushu, Japan. Journal of Geophysical Research: Solid Earth, 116(B2). https://doi.org/10.1029/2010JB007918.

Geyin, M.; B.W. Maurer; K. Christofferson; 2022, An AI driven, mechanistically grounded geospatial liquefaction model for rapid response and scenario planning, Soil Dynamics and Earthquake Engineering, Volume 159, 2022, 107348, ISSN 0267-7261, https://doi.org/10.1016/j.soildyn.2022.107348.

Goldberger, J.; S. Roweis, G. Hinton, and R. Salakhutdinov. Neighbourhood components analysis. In L. K. Saul, Y. Weiss, and L. Bottou, editors, Advances in Neural Information Processing Systems 17, pages 513–520, Cambridge, MA, 2005. MIT Press.

Green RA, Bommer JJ, 2019, What is the smallest earthquake magnitude that needs to be considered in assessing liquefaction hazard? Earthquake Spectra, 35(3):1441–1464. Russell A. Green, Julian J. Bommer; What is the Smallest Earthquake Magnitude that Needs to be Considered in Assessing Liquefaction Hazard?. Earthquake Spectra 2019;; 35 (3): 1441–1464. doi: https://doi.org/10.1193/032218EQS064M.

Guo, Y., T. Hastie, and R. Tibshirani. "Regularized linear discriminant analysis and its application in microarrays." Biostatistics, Vol. 8, No. 1, pp. 86–100, 2007.

Guyon, I., Elisseeff, A., 2003, An introduction to variable and feature selection, Journal of Machine Learning Research, 3, 1157–1182.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer Series in Statistics. New York, NY: Springer, 2009. https://doi.org/10.1007/978-0-387-84858-7.

He H, Bai Y, Garcia EA, Li S, 2008, ADASYN: adaptive synthetic sampling approach for imbalanced learning. In: 2008 international joint conference on neural networks (IJCNN 2008).

Hemez, Eds., pp. 27–43, Los Alamos National Laboratory, Los Alamos, NM, USA, 2005.

Hussain, J. N., 2008. Sensitivity analysis to select the most influential risk factors in a logistic regression model. Journal of Quality and Reliability Engineering. https://doi.org/10.1155/2008/471607.

James, Gareth • Daniela Witten • Trevor Hastie Robert Tibshirani; An Introduction to Statistical Learning with Applications in R, Springer New York Heidelberg Dordrecht London, 2013, ISBN 978-1-4614-7138-7 (eBook), https://doi.org/10.1007/978-1-4614-7138-7.

Jena, R., B. Pradhan, M. Almazroui, M. Assiri, H-J. Park, Earthquake-induced liquefaction hazard mapping at national-scale in Australia using deep learning techniques, Geoscience Frontiers (2022), https://doi.org/10.1016/j.gsf.2022.101460.

Knudsen, K., and J. Bott, 2011. Geologic and geomorphic evaluation of liquefaction case histories for rapid hazard mapping, Seismol. Res. Lett. 82, 334.

Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the fourteenth international joint conference on artificial intelligence; 1995. p. 1137–43. San Mateo, CA.

Kotoda K, Wakamatsu K, Midorikawa S. Seismic microzoning on soil liquefaction potential based on geomorphological land classification. Soils Found 1986; 28(2):127–43. https://doi.org/10.3208/sandf1972.28.2_127.

Li, P., Vu, Q.D. Identification of parameter correlations for parameter estimation in dynamic biological models. BMC Syst Biol 7, 91 (2013). https://doi.org/10.1186/1752-0509-7-91.

Liang L, Figueroa JL, Saada AS. Liquefaction under random loading: unit energy approach. J. Geotech. Eng. 1995; 121(11):776–81. https://doi.org/10.1061/(ASCE)0733-9410(1995)121:11(776).

Liu, D., & Zhang, H. (2018). Residuals and diagnostics for ordinal regression models: a surrogate approach. Journal of the American Statistical Association, 113(522), 845-854.

Manning, C. D., P. Raghavan, and M. Schütze. Introduction to Information Retrieval, NY: Cambridge University Press, 2008.

Marelli, S., & Sudret, B. (2014). UQLab: A framework for uncertainty quantification in Matlab (pp. 2554-2563). American Society of Civil Engineers.

Marelli, S., Lamas, C., Konakli, K., Mylonas, C., Wiederkehr, P., Sudret, B., 2022. UQLab user manual – Sensitivity analysis, Report UQLab-V2.0-106, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, 2022

Matsuoka, M., K. Wakamatsu, M. Hashimoto, S. Senna, and S. Midorikawa. 2015. "Evaluation of liquefaction potential for large areas based on geomorphologic classification." Earthquake Spectra 31 (4): 2375–2395. https://doi.org/10.1193/072313EQS211M.

Midorikawa S, Wakamatsu K. Intensity of earthquake ground motion at liquefied sites. Soils Found 1988; 28(2):73–84. https://doi.org/10.3208/sandf1972.28.2_73.

Oommen, T., L. G. Baise, and R. M. Vogel (2011). Sampling bias and class imbalance in maximum likelihood logistic regression, Math. Geosci. 43, 99–120.

Rashidian, V.; Baise, L.G.; Koch, M. Using High Resolution Optical Imagery to Detect Earthquake-Induced Liquefaction: The 2011 Christchurch Earthquake. Remote Sens. 2020, 12, 377. https://doi.org/10.3390/rs12030377.

Rashidian, V., and L. G. Baise (2020). Regional efficacy of a global geospatial liquefaction model, Eng. Geol. 272, 105644. https://doi.org/10.1016/j.enggeo.2020.105644.

Rokach L, Maimon O. Data mining with decision trees: theory and applications. World Scientific Pub Co Inc.; 2008, ISBN 978-9812771711.

Saltelli, A., Ratto, M., Tarantola, S., and Campolongo, F., 2005. "Sensitivity analysis for chemical models," Chemical Reviews, vol. 105, no. 7, pp. 2811–2827, 2005.

Saltelli, "Global sensitivity analysis: an introduction," in Sensitivity Analysis of Model Output, K. M. Hanson and F. M.

Seed, H. B., and Idriss, I. M. (1971). ''Simplified procedure for evaluating soil liquefaction potential.'' J. Geotech. Engrg. Div., ASCE, 97(9), 1249–1273.

Seyhan, E., & Stewart, J. P. (2014). Semi-empirical nonlinear site amplification from NGA-West2 data and simulations. Earthquake Spectra, 30(3), 1241-1256.

Shalev-Shwartz, Shai; Shai Ben-David, Understanding Machine Learning: From Theory to Algorithms, pp. 212 – 218; Cambridge University Press, 2014; https://doi.org/10.1017/CBO9781107298019.019.

Taborda, R., & Bielak, J. (2011). Large-scale earthquake simulation: computational seismology and complex engineering systems. Computing in Science & Engineering, 13(4), 14-27.

Terzaghi, K., Peck, R.B. and Mesri, G. (1996) Soil Mechanics in Engineering Practice. 3rd Edition, John Wiley and Sons, Inc., New York.

Theodoridis, Sergios, and Konstantinos Koutroumbas. Pattern Recognition. San Diego: Academic Press, 1999: 341-342.

Thompson, M., & Warmink, J. J. (2016). Natural hazard modeling and uncertainty analysis. Natural hazard uncertainty assessment: Modeling and decision support, 9-19.

Thompson, E. M., D. J. Wald, K. Allstadt, and M. Hearne (2016-b). Combining case history observations with different completeness levels in empirical ground-failure models, Seismol. Res. Lett. 87, no. 2B, 578.

Todorovic, Lana; Vitor Silva, 2022, A liquefaction occurrence model for regional analysis, Soil Dynamics and Earthquake Engineering, Volume 161, 2022, 107430, ISSN 0267-7261, https://doi.org/10.1016/j.soildyn.2022.107430.

Tokimatsu, K., Tamura, S., Suzuki, H., and Katsumata, K. (2012). Building damage associated with geotechnical problems in the 2011 Tohoku Pacific Earthquake. Soils and Foundations, 52(5):956–974.

Ujjwal, K. C., Garg, S., Hilton, J., & Aryal, J. (2020). A cloud-based framework for sensitivity analysis of natural hazard models. Environmental Modelling & Software, 134, 104800.

UNDRR: Global Assessment Report 2015, Making Development Sustainable: The Future of Disaster Risk Management, UN Office for Disaster Risk Reduction, Geneva, 2015a.

Van Ballegooy, S., Malan, P., Lacrosse, V., Jacka, M. E., Cubrinovski, M., Bray, J. D., ... & Cowan, H. (2014). Assessment of liquefaction-induced land damage for residential Christchurch. Earthquake spectra, 30(1), 31-55.

Vapnik V. The nature of statistical learning theory. New York: Springer; 1995.

Weiss GM, Provost F. In: The effects of class imbalance and training data size on classifier learning: an empirical study; 2001. https://doi.org/10.7282/t3-vpfwsf95.SN.

Worden CB, Thompson EM, Hearne M and Wald DJ (2020) ShakeMap Manual Online: Technical Manual, User's Guide, and Software Guide. Reston, VA: US Geological Survey.

Yang, W., Wang, K., Zuo, W., 2012, Neighborhood Component Feature Selection for High-Dimensional Data, Journal of Computers. 7 (1).

Youd, T. L., and D. M. Perkins (1978). Mapping liquefaction-induced ground failure potential, J. Geotech. Eng. Div. 104, no. 4, Paper Number 13659, 433–446.

Zhu, J.; Daley, D.; Baise, L.; Thompson, E.; Wald, D.; Knudsen, K., 2015, A Geospatial Liquefaction Model for Rapid Response and Loss Estimation. Earthquake Spectra 2015, 31, 1813–1837.

Zhu J., L. G. Baise, and E. M. Thompson (2017). An updated geospatial liquefaction model for global application, Bull. Seismol. Soc. Am. 107, 1365–1385. https://doi.org/10.1785/0120160198.

## Project Data

Each of the three projects are written up in peer-reviewed journal articles and Section 2 and 3 are part of a Ph.D. Dissertation by Adel Asadi published at Tufts. The peer-reviewed journal articles are in peer-review and electronic supplements will be provided with the journal article. The datasets used were published by others and all model parameters are included in the peer-reviewed publications.

## Bibliography

Adel Asadi, Laurie G. Baise, Magaly Koch, Babak Moaveni, Snehamoy Chatterjee, Yusupujiang Aimaiti (2023). Pixel-based Classification Method for Earthquake-Induced Landslide Mapping Using Imagery, Geospatial Data and Temporal Change Information. Under review by Natural Hazards Journal.

Adel Asadi, Laurie Gaskins Baise, Snehamoy Chatterjee, Weiwei Zhan, Alexander Chansky, Babak Moaveni (2023). Regionally Informed Global Geospatial Modeling of Earthquake-Induced Soil Liquefaction Using a System of Voting Machine Learning Classifiers. In preparation for submission to Engineering Geology Journal.

Zhan, W., L.G. Baise, B. Moaveni (2023). An Uncertainty Quantificaiton Framework for Logistic Regression based Geospatial Natural Hazard Modeling. Under review by Engineering Geology Journal.