PREDICTING PIPING EROSION SUSCEPTIBILITY BY STATISTICAL AND ARTIFICIAL INTELLIGENCE APPROACHES- A REVIEW

Crissa Meriam George[1], Anu VV [2]

[1]PG student, Civil Engineering Department, Toc H Institute of Science and Technology, Kerala, India
[2]Assistant Professor, Civil Engineering Department, Toc H Institute of Science and Technology, Kerala, India

---

Abstract - *Internal erosion is a general term referring to the mechanism of detachment and movement of soil grains due to water flow through a porous media. It includes many different processes such as piping, soil contact erosion, or suffusion. The study and prediction of these complex phenomena is a recurrent topic in many different fields. This is because internal erosion is one of the main causes of failure of water retaining structures such as dikes and dams. Piping is a complicated natural phenomenon of internal erosion, which threatens the safety of hydraulic earth structures. The objective of this paper is to develop an approach by using three machine learning algorithms—mixture discriminant analysis (MDA), flexible discriminant analysis (FDA), and support vector machine (SVM) in addition to an unmanned aerial vehicle (UAV) images to get map susceptibility to piping erosion. By performing this approach it is expected to obtain an efficient piping susceptibility map can be prepared.*

***Key Words**: Soil piping, Machine learning models- mixture discriminant analysis (MDA), flexible discriminant analysis (FDA), and support vector machine (SVM).

1. INTRODUCTION

Many field observations have led to speculation on the role of piping in embankment failures, landslides and gully erosion. Piping, defined as the movement of fine particles out of the soil layer by the force of seepage through the soil is one of the most common modes of failure in embankment dams and foundations. The internal potential for soil piping is mainly controlled by the grain size distribution and the skeleton structure of the soil. The significance of piping includes its link to gully development and direct land degradation processes. It plays an essential role in storm runoff, loss of agricultural productive capacity, environmental problems in downstream areas, and increased sediment yields. The effect of piping is seen in ecosystems and their abiotic factors where considerable disturbances to ecosystems can impact the availability of their services and functions. Piping can occur in different climates and various soil types. Numerous researches reported the impact of both physical and chemical soil properties on piping, as well as the impact that dispersive properties of soils have on pipes. The factors that control pipe formation are often related to soil properties. There is a high probability of that pipes will appear and develop in sodic soils or non-sodic soils, and with high percentages of exchangeable sodium (Na), high contents of soluble salts, organic soils, lithologic discontinuities in soils, silt-rich soils, texture-contrast soils, dispersive soils, and soils that have macrospores of different origins. Further, there is a high probability that pipes will appear and develop in a variety of soil types and compositions, including those prone to erosion.

1.1 Soil piping

Soil piping erosion or tunnel erosion is one of the factors which lead to formation of land subsidence. It is defined as the hydraulic removal of subsurface soil, causing the formation of underground channels and cavities. This is not a Universal process but it is important to certain localities. The evidences shows that piping performs as a function of drainage and erosion but the drastic collapse of roof of the soil strata makes it one of the silent triggers. Piping has been observed in both natural and anthropogenic landscapes, in a wide range of geomorphologic, climatological and pedological settings. Rainfall and excess run off after the initiation of the piping erosion entrain more dispersed clay particles, resulting in both head ward and tailward expansion of cavities until a continuous pipe is. And at the final stage piping may reach to an extent where complete roof collapse occurs and erosion gullies form. The geophysical and geochemical factors control the process of soil piping. The geophysical parameters include slope gradient, drainage, type of soil (clay, silt, sand, pebble, cobble, boulders), soil texture (sand silt, clay %), infiltration rate, amount of rainfall, porosity, permeability, dense of vegetation etc. The geochemical parameters include clay mineralogy, pH of soil, conductivity and total dissolved salts.

Land subsidence by piping: Land subsidence is a gradual settling or sudden sinking of Earth's surface due to movement of earth materials. In highlands of Kerala during monsoon season land subsidence has become a very common phenomenon which is a threat to human life in all aspects. Land subsidence occurs naturally and artificially. Natural subsidence occurs due to many factors. Land subsidence can occur over a large area than a small like sink hole. Subsidence is a global problem that occur mainly because of exploration of underground water increasing development of land and water resources threatens to exacerbate existing land subsidence problem and initiate new one. Loss of subsurface support is found to be the main reason for subsidence to occur. Water percolating through pervious surficial materials gets diverted to weak pervious portions and the dispersive soils also carries out with this water leading to huge cavities.

Causes of land subsidence: Lowering of the land surface of large areas has been a major unintended consequence of ground water and petroleum withdrawal by humans.

Land subsidence causes many problems including:

- Changes in elevation and slope streams, canals and drains
- Damage to bridges, roads, railroads, storm drains, sanitary sewers, canals, and levees
- Damage to private and public buildings
- Failure of well casings from forces generated by compaction of fine-grained material in aquifer systems.
- Permanent inundation of land, aggravates flooding, changes topographic gradients and ruptures the land surface.
- Reduces the capacity of aquifers to store water. Due to its ability to destroy property on a large scale, subsidence is a very expensive type of mass wasting that also poses some risk to human lives.

1.2 Machine learning model

Machine learning models are useful to effectively identify zones susceptible to piping, the spatial occurrence of piping erosion, and its controlling factors. These models handle data from different measurement scales and work with various kinds of independent variables. Three machine learning algorithms for spatial modeling of piping are considered they are namely, the support vector machine (SVM), flexible discriminant analysis (FDA), mixture discriminant analysis (MDA), and logistic regression (LR) models were used to predict pipe susceptibility maps in an area highly prone to piping erosion. Objectives are to:

- Spatially predict the occurrence of piping erosion
- Develop SVM, MDA, and FDA machine learning models to predict the potential for piping and compare these results with the LR data mining model

Use various sample data sets and evaluation criteria to assess the capability and robustness of these machine learning models.

The methodology consisted of five phases:

1. Selection of the study area
2. Data preparation that included construction of the piping inventory map and conditioning factors [field investigation, acquiring aerial photos by an UAV images, and laboratory soil analysis
3. Data correlation analysis with ME (maximum entropy);
4. Piping spatial modelling using SVM, MDA, FDA, and LR
5. Validation of models.

Description of the study area include: Location of the area susceptible to erosion in terms of longitude and latitude, estimate area(ha) and the mean slope of the region, calculate the annual rainfall(mm) and average temperature(°c), the types of soil contained study area has to be surveyed, the source of parent rock from which soil is derived and land use and geology of the study area should be found. After determining the study area, we perform extensive field surveys in the sub-catchment. Area scrolling is conducted to locate and measure the types of piping forms that lowered the soil surface without any break in the vegetation cover and hollows with more or less vertical walls that connected to the pipes. Further, field surveys have to be conducted to identify piping collapses and conditioning factors, then use UAV system to acquire aerial photographs.

In addition to the piping inventory map, various inter-related factors also affect the pipes. Conditioning factors such as altitude, slope length (LS), slope aspect, slope degree, topographic wetness index (TWI), land use, stream density, distance from streams, distance from roads, terrain ruggedness index (TRI), convergence index (CI), profile curvature, plan curvature, silt, clay, sand, pH, exchangeable sodium percentage (ESP), soil weight in a given volume (bulk density), soil electrical conductivity (EC), calcium (Ca+2), magnesium (Mg+2), and sodium (Na) for the spatial modelling and mapping of the pipes. These factors are selected according to data availability, interviews with farmers and local stakeholder, field investigation. However, there is no universal guideline for selecting the parameters for piping susceptibility assessment. All of the data layers were prepared in raster format with a spatial resolution of 1 m or a pixel size of 1 × 1 m2. Altitude, slope aspect, slope degree, TWI, TRI, and CI were derived from the 1 m resolution digital elevation model (DEM). Distances from the roads and streams, and land use are obtained from the aerial photos.

Support vector machine (SVM): Based on statistical learning theory, SVM follows the principle of structural risk minimization and is used as a machine learning algorithm. The two main principles of SVM are: the optimal classification hyperplane and the use of a kernel function. The dots and squares represent two types of samples, H is the classification line between them, and H1 and H2 are lines parallel to H running through the sample points closest to the classification line (the support vectors). The distance between them is called the classification margin. The goal of the optimal classification hyperplane is to discriminate between the two types of samples correctly (though certain errors are allowed) while maximizing the classification margin.

Mixture discriminant analysis (MDA): The MDA model successfully combines the performance characteristics of more complex neural network models, which is attributed to the nonlinear nature of its classification rules along with the ease of interpretation associated with linear mixture models partly due to its relatively simple structure. These models provide a good alternative to assess the problems with mixture modelling in remote sensing.

Flexible discriminant analysis (FDA): This method is a generalization of linear discriminant analysis (LDA) to a non-parametric version. FDA can be used for post-processing a multi-response regression that employs an optimal scoring technique. The goal of FDA is correspondence between LDA, canonical correlation analysis, and optimal scoring. FDA can be performed as a multi-response linear regression that uses optimal scoring to represent the classes. It is equal to conducting an LDA in the space of fitted values from the regression of the class indicators on the predictors. This method is a substitute for the linear regression step by a non-parametric or semi-parametric regression, which paves the way for a variety of regression tools to provide various discrimination rules and flexible class boundaries.

Logistic regression (LR): The LR is a multivariate statistical regression analysis. The model has been widely applied for LS mapping. LR is a type of multivariate regression. It is utilized to find the relationship between a dichotomous response variable, which is coded as 0 (corresponds to the absence) and 1 (corresponds to the presence), and a set of explanatory variables x1, x2,....xn both categorical and numerical. The LR model provides an explanation of the relationship between the dichotomous response variable, as the presence or absence of a collapsed pipe, and a set of independent explanatory variables. With dummy coding, each parameter estimates the difference between that level and the reference group.

Validation Process: The performance of the used models are evaluated using a standard technique namely, the Receiver Operating Characteristic (ROC), which has been applied in the studies of geo-hazard modeling. Pipes that were not used in the training process (30%) were used to validate the predictive capabilities of the four models, as verification of the piping susceptibility maps. Receiver Operating Characteristic (ROC) curves are one of the most commonly-used performance techniques which are applied to evaluate the overall performances of the models. The ROC curve is a beneficial tool to represent the quality of deterministic and probabilistic forecasting systems, in which the sensitivity (the true positive rate) of the model is the y-axis and the 1-specificity (the false positive rate) is the x-axis. An ROC curve is better than another because of its cutoff-independent. Also, the operating range of values is better than a model which classifies objects by chance. The AUC (Area under the curve) is calculated by Equation.

$$\mathrm{AUROC} = \frac{\mathrm{aTP} + \mathrm{aTN}}{\mathrm{P} + \mathrm{N}}$$

where TN and TP are the percentage of negative and positive instances which are correctly classified, P is the total number of pipes and N is the total number of non-pipes.

Higher AUC values indicate a better model when the AUC ranges from 0.5 to 1.0.

Table-1: Prediction accuracy based on AUC values

AUC values	Prediction accuracy
0.9-1	Excellent
0.8-0.9	Very good
0.7-0.8	Good
0.6-0.7	Average
0.5-0.6	Poor

Variable importance analysis: Using ME and LR models, participation rate of variables and their importance are determined. The ME model is specifically designed for ecological modelling and species distribution assessment. The principle of maximum entropy (ME) states that the probability distribution which best represents the current state of knowledge is the one with largest entropy. According to this principle, the distribution with maximal information entropy is the best choice. All input conditioning factors are introduced as random variables of the model according to the ME algorithm, which represents their uncertainty. In order to assess the uncertainty of predicting the piping potential according to the ME model, a jack-knife test can be conducted to examine the effects of removal of any of the conditioning factors on the potential map. This test gives access to contributions by the factors (i.e., relative importance).

Validation: UAV images reduces the time and cost of geomorphologic mapping because of their ability to acquire images of high spatial resolution. Thereby increasing the precision of the acquired dataset and reducing time for field-related activities, it enables the detailed mapping of different erosion types. This is found to be an attractive tool to monitor and map different aspects of the environment. The results of the four models provide different ranges of susceptibility values for the piping process. So we obtain four susceptibility maps produced by SVM, MDA, FDA and LR models. The susceptibility classes for each model are defined as high, moderate and low. The area percentages in each class of models are prepare

2. CASE STUDY

Golestan Province is located in Northeast Iran on the south-eastern shore of the Caspian Sea. The study area, the Iky Aghzly sub catchment, is part of the Gorganrood Catchment in Golestan Province that is located between 55° 38′ to 55° 40′ Eastern longitude and 37° 37′ to 37° 39′ Northern latitude. This area comprises a region of 703.36 ha with a mean slope of 27%. The Iky Aghzly sub catchment represents a hilly area with altitudes that range from 336.374 to 548.01 m a.s.l. According to the Iranian Meteorological Organization, there is an average annual rainfall of 385 mm and an average temperature of approximately 18.2 °C. The study area has a silty surface soil layer which has been derived from loess deposited. It consists of two soil textures, silt-clay-loam and silt-loam. The stair, circle, ellipsoid, diamond, triangle, and rectangle piping forms have been seen in this area. Land use in the study area includes agriculture and rangeland. Triticum is the primary agriculture crop, whereas the rangeland is mostly covered with grasses, Paliurus spina-christi, Phragmites australis, Punica granatum, and Ficus carica. In this area, we found 42 (~12%) pipes in the agricultural lands and 303 pipes (~88%) in the rangelands. Geologically, the study area is part of the Alborz zone of Iran and consists of Miocene sediments including calcareous sandstone,sandy limestone with marl and conglomerate.

2.1 Piping susceptibility map

In this study, the geomorphic piping phenomenon in a loess-covered region of Northeast Iran. UAV were used to reduce the time and cost of geomorphological mapping because of their ability to acquire images at very high spatial resolution for precise recognition of features. The aerial photos from UAV photogrammetry could easily detect small geomorphic units. This technique has a significant effect in reducing the time for field-related activities and allows for increased precision of the acquired dataset.

The high resolution images captured by UAV are valuable datasets that enable detailed mapping of different erosion types. In combination with ground thrusting images, the UAV images provide a reliable tool for estimation and recognition of the relationship between piping processes and a set of independent explanatory variables. The results of the four models provided different ranges of susceptibility values for the piping processes. The training data is used to construct the SVM model, which was standardized to 0 and 1. The four susceptibility classes for each model were defined as very high, high, moderate, and low. The SVM model results indicated that the low class had the largest area (46.35%), followed by the moderate (29.33%), high (14.07%), and very high (10.25%) classes. The results on piping susceptibility zoning according to the MDA model showed that the low class also had the largest area at 39.81%, followed by 31.58% (moderate), 19.25% (high), and 9.36% (very high). In the FDA model, the highest area was seen in the high class at 34.38%, followed by the moderate (27.6%), low (20.7%), and very high (17.23%) classes. In the LR mode, the low class has the largest area (81.73%), followed by the moderate (5.62%), high (4.59%), and very high (8.06%) classes.

2.2 Validation of the piping susceptibility maps

The results of the training step confirmed that the SVM model had the highest AUC value (93.8%). The other models had an AUC of 92.4% (MDA), 90.9% (FDA), and 90.3% (LR). The results of the validation step showed that the SVM model had the highest AUC value (92.45%). The AUC for the prediction-rate curve produced by these models was 91.34% (MDA), 90.32% (FDA), and 89.27% (LR). Based on the results, the SVM model had the best AUC (92.45%). Therefore, it can be concluded that the SVM, MDA, and FDA machine learning models were excellent with AUC values>90%. However, the LR model was very good with the AUC values> 80 The SVM model also had the best capability to predict piping susceptibility, followed by the MDA and FDA models. They confirmed that pipes were more likely to happen when a topographical threshold that relied on both slope gradient and upslope area was exceeded in areas that had adequate water supply from topographical convergence.

2.3 Variable importance analysis

According to the results, silt content had the greatest impact on the occurrence of piping, followed by TWI, upslope drainage area, bulk density, profile curvature, and land use. Although there is a lack of agreement about the factors that contribute to subsurface pipe development and the interactions of pipes with these factors, a number of researchers have confirmed the effect of silt content on piping activity. Patterns of collapsed pipes are highly related to topohydrologic conditions. These highly susceptible areas are usually places where there is a likelihood for increased water concentration and subsoil saturation. The soil properties of their study area clearly differed from the loess derived soils in the current study.

3. CONCLUSION

The susceptibility to piping is influenced by a set of geo environmental conditions. This study aimed to assess the geomorphologic role of piping and the factors that control pipe formation. The effectiveness of the machine learning algorithms SVM, MDA, FDA, and LR in predicting areas that had a high probability for piping. The accuracy of the model are tested with the ROC curve. It is found that the curve had accuracy with tested by combining all the four machine models.According to the mechanism of the landslide, certain factors which are not involved in the analysis can have significant impact on landslide occurrence. Underground water can have a significant effect on some sites while it is generally not considered in landslide susceptibility assessment. For rock slopes, weathering plays an important role in landslide development, but this too is not taken into account. A more comprehensive database should be established using multi-source and multi-scale monitoring technology. It is important to distinguish the type of landslide when using machine learning methods to assess landslide susceptibility. More data about the landslide type should be included. These models were applied because they can model the non-linear relationship between piping erosion occurrence and its conditioning factors. The models present cost effective and accurate results. The main advantage of this model is that the model gives a ranking based on the susceptibility of landslide in that particular area more precisely. But they cannot quantify pipes development and their temporal changes.

REFERENCES

[1] Mohsen Hosseinalizadeh, “How can statistical and artificial intelligence approaches predict piping erosion susceptibility?” Science of the Total Environment, vol 646, page no: 1554–1566

[2] Alba Yerroa, “Modelling internal erosion with the material point method”, 1st International Conference on the Material Point Method, vol 175, page no:365 – 372

[3] Jiansheng Chen, “Experimental Investigation Of The Erosion Mechanisms Of Piping”, Soil Mechanics and Foundation Engineering Vol. 52, page no:5.

[4] Yang Yang, “ An Experimental Investigation of Piping Effects on the Mechanical Properties of Toyoura Sand”, Geotechnical engineering page no:1-10

[5] Anita Bernatek-Jakiel, “Assessment of grass root effects on soil piping in sandy soils using the pinhole test”, Geomorphology, vol 295, page no:563–571

[6] Aril Aditian, “Comparisonof GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia”, Engineering Geology,

[7] AnitaBernatek-Jakiel, “ Subsurface erosion by soil piping: significance and research needs”, Earth-Science Reviews, vol 185, page no:1107-1128

[8] Binh ThaiPham, “A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area” (India), Geotechnical engineering.

[9] Jia-JyunDonga, “Discriminant analysis of the geomorphic characteristics and stability of landslide dams”, Geomorphology.

[10] SanweiHeab, “Application of kernel-based Fisher discriminant analysis to map landslide susceptibility in the Qinggan River delta, Three Gorges, China”, Geotechnical engineering.

[11] Gwo-Fong Lina, “ Assessment of susceptibility to rainfall-induced landslides using improved self-organizing linear output map, support vector machine, and logistic regression”, Engineering geology, vol.224, page no:67-74.

[12] Arsalan Ahmed Othman, “Improving landslide susceptibility mapping using morphometric features in the Mawat area, Kurdistan Region, NE Iraq: Comparison of different statistical models”, Remote sensing of Environment.

[13] Andrea Ciampalini, “Landslide susceptibility map refinement using PSInSAR data”, Remote Sensing of Environment, vol 84, page no: 302–315

[14] Jente Broeckx, “A data-based landslide susceptibility map of Africa”, Remote Sensing of Environment.

[15] Ephemeral G.V. Wilson , “, Gully erosion by preferential flow through a discontinuous soil-pipe”, USDA-ARS National Sedimentation Laboratory, Vol 73, page no:98–106