

Data Mining Applied for Liquefaction Mapping and Prediction Learn from Palu Earthquakes

Andri Irfan Rifai^{1,*}, Hendra², Eko Prasetyo³

¹Faculty of Engineering & Planning, Universitas Internasional Batam, Indonesia

²Faculty of Engineering, Universitas Siliwangi, Indonesia

³Faculty of Engineering, Universitas Tadulako, Indonesia

Received February 25, 2020; Revised June 24, 2020; Accepted July 7, 2020

Copyright©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract The 2018 Palu-Sigi-Donggala earthquakes in Center Celebes have caused significant damage to many residential houses due to varying degrees of soil liquefaction over a vast extent of urban areas unseen in past destructive earthquakes. Soil liquefaction occurred in Palu and Sigi, thus providing researchers with a wide range of characterizing soil and site response to large-scale earthquake shaking. One of the essential learning issues is the prediction of liquefaction. Prediction of liquefaction is also a complex problem as it depends on many different physical factors, and the relations between these factors are highly non-linear and complex. Most of these approaches are based on classical statistical criteria and neural networks. In this paper, a new method which is based on classification data mining (DM) is proposed. The proposed approach is based on historical data from the field and sciences portal. The proposed algorithm is also compared with several other DM algorithms based on the miner. It is shown that the proposed algorithm is very useful and accurate in the prediction of liquefaction.

Keywords Liquefaction, Data Mining, Prediction

1. Introduction

Celebes Island is one of the largest islands that spread in the Indonesian archipelago. The K-shaped island is in the eastern part of Indonesia, which is converge of Eurasia, Indo Australia, and Pacific tectonic plates. The third meeting of the plates led to very active subduction and collision, which caused quite complex problems.

The plates in the Southeast Asian region are the result of

thousands of years of tectonic evolution. Tectonic activities take place and are centred in Indonesia, Malaysia, New Guinea, and northern Australia (Hall & Spakman, 2015). The geological area, which was later formed, still leaves conditions that are still very interesting to study. Various technical activities continue. Also, the volcanic activity does not want to be left behind as can be seen in Figure 1 which shows the volcanic distribution that extends from Sumatra Island, Sunda Strait, Java Island, Bali Island, Nusa Tenggara, Celebes. It continues to the Islands in the Philippine State Region.

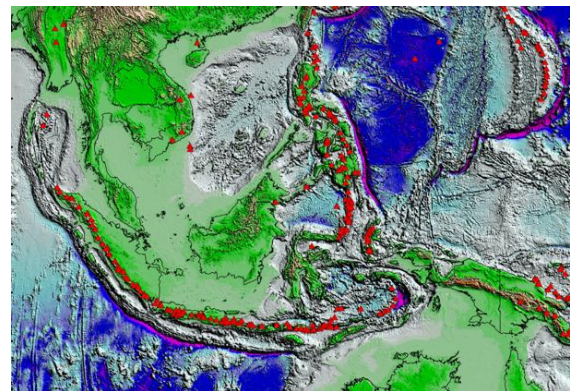


Figure 1. Distribution of three main plates (Hall & Spakman, 2015)

The volcanic and tectonic activities in the Sulawesi Island area encourage plate movement activity combined with the Koro fault making this area very active (Thein et al., 2015). From searches on the USGS portal sciences in the last 100 years, there have been 104 earthquakes with magnitudes > 5. The distribution of regions that are not too broad, as shown in Figure 2 causes the intensity of influence received by one point is quite significant.

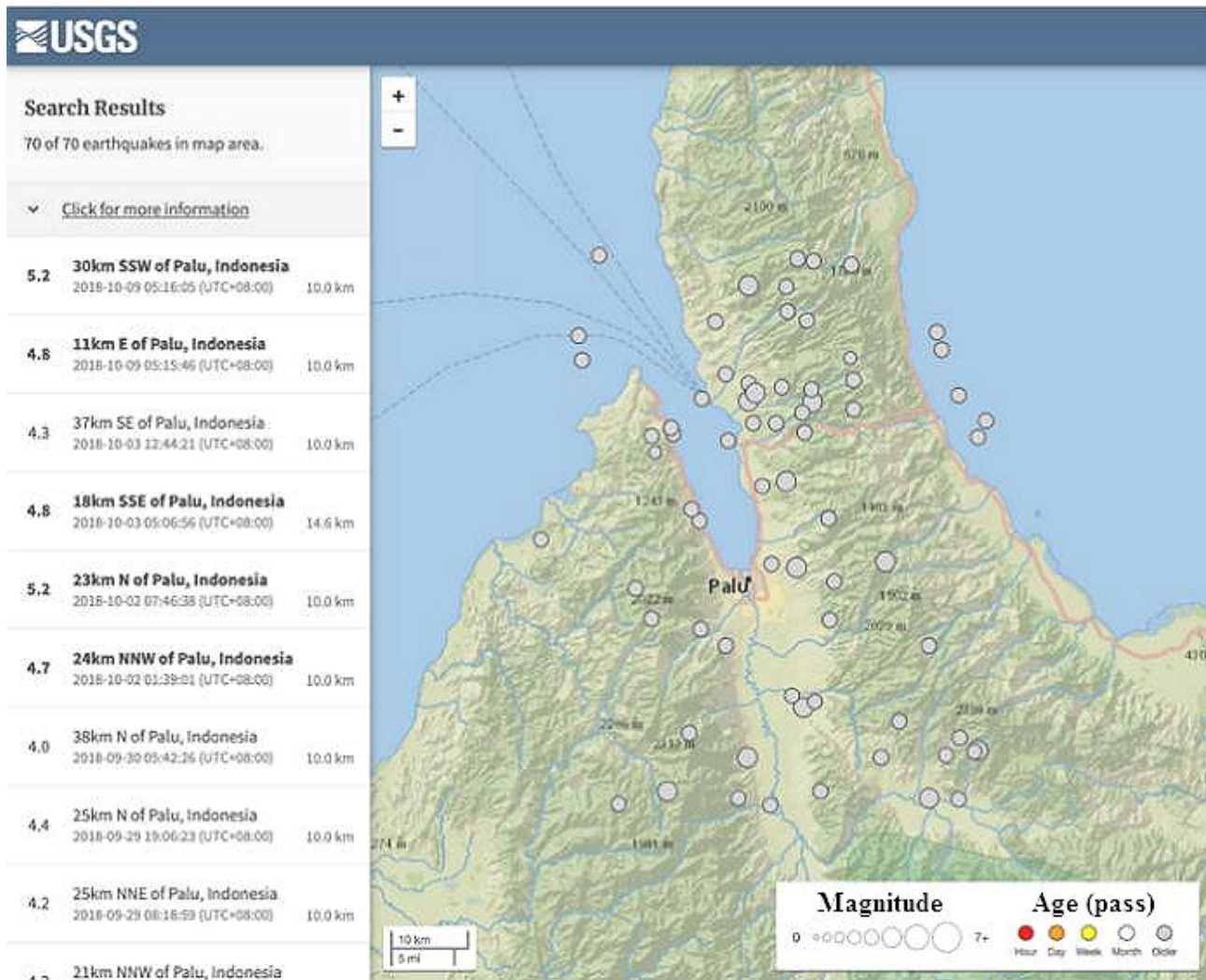


Figure 2. Earthquakes in the 100-last year (USGS, 2019)

From several earthquakes, if grouped into the largest scale, it can be seen some with characteristics that are interesting to study, i.e.

1. It happened on September 28, 2018, at 10:02:45 (UTC); with the location 0.256°S $119,846^{\circ}\text{E}$; Scale of 7.5; with a depth of 20 km.
2. It happened on October 10, 1998, at 16:32:19 (UTC); with locations 0.403°S $119,840^{\circ}\text{E}$; scale of 6.0 depth of 33 km
3. It happened on April 6, 1965 at 09:42:29 (UTC); with location 0.463°S 120.039°E ; scale 6.3; depth of 25 km.

The earthquake that occurred on September 28 was quite shocking to all quarters because the earthquake was accompanied by tsunami waves and liquefaction with the loss of lives of thousands of people. The effect of the

earthquake was quite large because the epicentre was not too deep, and the frequency of the earthquake was quite intense. In figure 3, the distribution of earthquakes can be seen. Serva et al., (2016) states that earthquake intensity that occurs can be measured through a scale by considering the incidence, size, and wide distribution of earthquake environmental effects. Also, one of the interesting things is the epicentre form of the earthquake that is formed is directed to the city of Palu.

The Palu region had the biggest influence when the earthquake occurred, aside from the quite high intensity of the earthquake it was also influenced by the presence of a Palu-Koro fault that bisected the area of the city. This fault is an active fault that extends from central Sulawesi to the Karimata strait (Mallick et al., 2018). Socquet et al., (2019) estimate this fault moves actively with a shift of 41-45 mm/year.

Figure 3. The instrumental intensity of Palu earthquakes (USGS, 2019)

Liquefaction that occurred shortly after the biggest earthquake shook the city of Palu still leaves a variety of interesting questions, ranging from the area affected and the form of liquefaction that also occurs quite varied. There are four areas with the greatest influence of liquefaction, namely Petobo, Balaroa, Jono Oge, and Sibalaya. In the standard theory, it is stated that liquefaction can occur if sand soil (sand, silty sand, clayey sand) which contains a little clay and graded tends to be uniform.

2. Literature Review

Several liquefaction predictions approaches have continued to develop in recent decades. One of them is Al Bawwab (2005) which began to group several approaches that can be used in predicting liquefaction events, namely:

1. Numerical analysis of finite elements;
2. Simplified analytical methods;
3. Soft computing methods;
4. Empirical methods were conducted based on field, laboratory, and statistical tests.

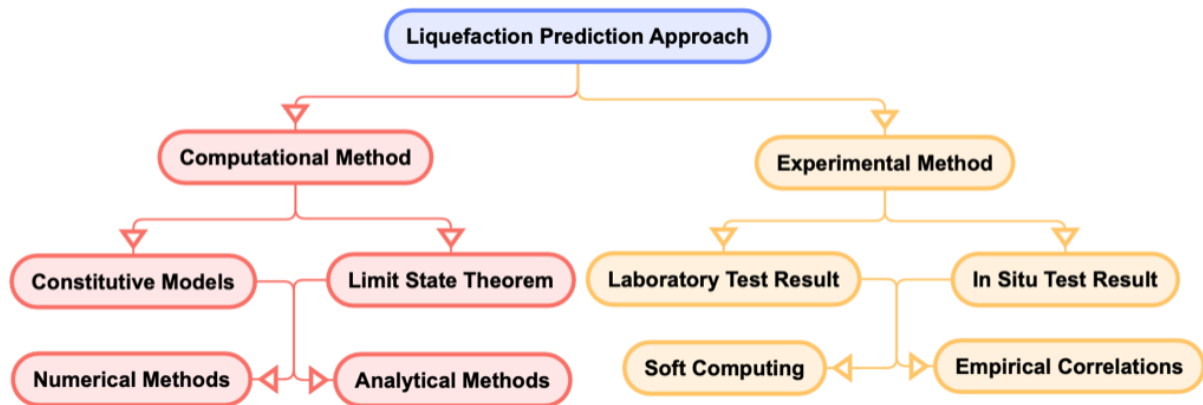


Figure 4. Classification of predictive approaches Earthquakes (Kalantary et al., 2013)

Subsequently, in 2013, Kalantary et al., simplified the various approaches above into diagrammatic form as can be seen in Figure 4.

Following the flow of thought of the concept approach presented in Figure 4, two fundamental approaches are explained, namely the computational approach and the experimental approach. The latter approach explained that the results of laboratory and field tests are used as a series of historical data that can be used to develop empirical correlations, meanwhile, in the method of computing the basic parameters incorporated into the analytical or numerical models to make predictions on the trends that occurred.

Numerical methods and analytical methods have been developed and used in calculating and predicting geotechnical phenomena. The success rate with this method is quite high, but requires quite complex calculation steps and requires repeated analysis. One method that can be used is the finite element method (Galavi, Petalas & Brinkgreve, 2013; Mohammadnejad & Andrade, 2015).

The dynamics of geotechnical characters that occur in liquefaction are quite high, so a method that is more accommodating than numerical analysis or the like is needed. In response to this, various soft computing and empirical correlation approaches began to be developed. Some of the approaches that have been carried out by previous researchers are models for assessing lateral spreading liquefaction. (Javadi, Rezaia, & Nezhad, 2006; Goh & Zhang 2014). Furthermore, the empirical model uses a lateral displacement index approach (Zhang, et al., 2004; Franke, et al., 2016).

Finally, the latest approach is a popular method in the form of fuzzy logic models, neuron computing, probabilistic reasoning, and genetic algorithms (Goharriz & Marandi, 2016). The fuzzy logic approach is then developed and sharpened by artificial neural networks (Shahri, 2016; Sulewska 2017). There was noting the existence of massive data and recording of earthquake and liquefaction phenomena lately. The latest technological approach is needed so that the collected data can be utilized

in a structured and measured manner to support a better risk mitigation system through accurate interpretation and prediction of data. DM is one of the latest approaches that are capable of making fairly accurate interpretations in predicting the latest technical trends (Rifai, et al., 2015).

After the process of data interpretation and trend, the prediction is done, and it is necessary to optimize mitigation from the beginning so that the success factors that influence it in the form of technical data can be recorded, archived, and evaluated. In this regard Rifai, et al. (2016) has developed an optimization model using information technology that is supported by a combination of mathematical approaches and genetic algorithm optimization. Although the approach is for other technical approaches, the concept of the approach can still be referred to as forming trends.

From the various descriptions of the development of the prediction method above, in this paper, a new approach is proposed that proposes the benefits of a combination of empirical and optimization methods by considering various independent variables that are integrated with geotechnical science.

3. Research Method

In this research, an approach is needed to integrate empirical data on the occurrence of liquefaction and in situ tests of soil conditions and the results of topographic measurements. The modelling objectives developed can make changes to the mitigation plan more accurate when needed, at the specified location and the type of handling that is intended to be applied interactively and can demonstrate its impact on the dynamics of geotechnical conditions. This allows the stakeholder to conduct analysis more effectively so that it can better support the mitigation process.

Data Entry dan Acquisition

Data is one important component in DM, and the

methods available to add or obtain data are the most important. The methods to achieve this are:

1. Import digital information available in compatible formats;
2. Using available data [one of which can use USGS];
3. Digitalization of analog data.

On the one hand, the development of information technology with the emergence of digital information and databases has made data access easier, especially those based on GIS, science data catalogues, geologic maps. Additionally, compatibility between software is increasingly common, which makes it possible to convert data originating in one software to be used in another format. Formats such as CAD (for example, DWG, DXF), vector and raster, widely used commercial GIS data (ARC/Info, ARC / View, MGE Intergraph, etc.), and general image data (for example, tiff, BMP, and others).

Besides, it is currently widely known that a method for acquiring geological map data is the use of DM-bridges. The method is carried out with various purposes, such as mapping, determining the coordinates of handling, determining the extent of events, and so on. The data used to simulate events is earthquake data in Palu and its surroundings, while for interface simulations using earthquake Palu-Koro fault mapping.

Mapping Implementation and Analysis

Among the various abilities of DM, neural analysis is an important stage in this research. The limited analysis is carried out on vector data, because their use can illustrate predictions of existing history, which are defined as a series of interconnected features, representing geological characteristics. Mapping, drawing, and presenting area coverage using conventional methods can cause inaccurate decision making. This open-source application used has the ability and facilities to integrate with various systems.

The information system application software used in this study was implemented from import data on the USGS science portal for a changing world, then using an advanced algorithm performed analysis.

Analysis and Prediction

In conducting analysis and prediction, the steps are through several steps, as follows:

1. Historical criteria. Information on earthquake events which are then followed by liquefaction events can be obtained from field data records, satellite data, or data provided on certain sciences portals. The liquefaction event is indeed related to the earthquake magnitude, but this is not a guarantee of certainty that the liquefaction will occur.
2. Geological criteria. The geological process that forms soils consists of uniform grain distribution and soil sediment in loose conditions having high liquefaction susceptibility. There were many parameters, such as

3. Composition criteria. Land with good gradation generally has a low potential for liquefaction. Grading criteria, composition, particle shape, and the like are as detailed as possible.
4. Deep learning prediction method. By using the learning concept approach from the available big data, the various possibilities for liquefaction are studied.

Data Mining Task

DM tasks are arranged based on the ability of DM to solve various problems with interpretation and other statistical operations on data (Freitas, 2013). Depending on the type of pattern found, DM tasks are usually classified into two categories, namely predictive and descriptive. Predictive approaches make inferences in the data to predict unknown values of the output variable, taking into account known values of the input variable (Wu et al., 2014), while a descriptive approach is to characterizing and summarizing the various general characteristics of the data to increase understanding and broad information. The ability of the DM task is very dependent on the user's ability to identify the initial problem and the purpose of its resolution.

R Tools for DM

R is a software unit integrated with several facilities for computational engineering with computational capabilities and reliable graphics performance. R is a group of programming languages-S, developed by AT&T Bell Laboratories (now Lucent Technologies) in the late 1970s. R is a free version of the S language software similar to S-PLUS, which is widely used by researchers and academics in scientific activities. The R environment is open source and a high-level matrix programming language, widely used for data analysis and statistics. R can perform statistical operations (linear and non-linear modelling, classical statistical tests, classification, clustering, etc.), graphics, object-oriented design, flexible programming. This application is also made extensible easily with the creation of new applications that can be implemented through package development. The R community is very active, and new packages are constantly being developed. So in this perspective, R can be open source to share algorithms throughout the world (Cortez, 2010). One such package is the r-miner available at <http://www3.dsi.uminho.pt/pcortez/rminer.html>, to facilitate the use of the DM algorithm to carry out classification and regression.

4. Discussion and Result

The results of research in compiling the concept of a

liquefaction prediction approach can be described as follows.

Liquefaction Potential Index

The Liquefaction Potential Index (LPI), which was developed starting in the 1970s by Iwasaki et al. (1984) to predict the potential for surface liquefaction, can be correlated directly with the potential damage and cumulative response of soil sediment. Iwasaki et al., Stated that LPI can be interpreted with very low liquefaction risk if $LPI = 0$; low if $0 < LPI < 5$; high if $5 < LPI < 15$; and very high if the $LPI > 15$. Also, the potential for liquefaction takes into account several things, namely:

1. The proportion of layer thickness that is liquefaction,
2. The proportion of the closeness of the layer that is liquefaction to the soil surface is related to the safety factor for the initiation of liquefaction, but only soils with $FS < 1$ have potential liquefaction.

Liquefaction Severity Index

Liquefaction Severity (S) is defined as a horizontal ground movement in inches. S began to be investigated by Youd & Perkins (1987) followed by several subsequent researchers, including Sonmez & Gokceoglu (2005). The study showed very little damage shown by ordinary buildings with a value of $S < 5$, but low to high damage has a value of $S = 5$ to 20 and major damage has a value of $S > 30$.

At present, the maximum S value for the lateral spread is developed due to a large flood area, delta, or other areas with a slope of fluvial Holocene sediment. Furthermore, it is simplified with a liquefaction severity index (LSI) which can be normalized by taking into account seismologic, geologic, topographic, hydrologic, and geotechnical factors. The final regression from the Youd and Perkins equation is as follows:

$$\text{Log (LSI)} = -3.49 - 1.86(\log) + 0.98 M_w \quad (1)$$

Where R is the epicentre distance in kilometres and M_w is the moment magnitude.

Liquefaction Displacement Index

Lateral displacement index (LDI) according to Zhang et al., (2004) and Faris et al., (2006) is the integration of lateral displacement through the approach of permanent shear strain to depth to calculate lateral spreading movements on the surface of the soil. Faris et al. (2006), uses LDI to develop equations for maximum horizontal movement from laterally spreading areas that are functions that are equivalent to the amount of blow in clean sand, slope, α , and moment magnitude, M.

Case Study: Palu Liquefaction 2018

Palu disaster shows the challenges facing the residents

are very large, because the tsunami and liquefaction arrived very quickly, in just a few minutes without any prior warning. This is very different from what happened in Japan (2011), which has a lot of time to give warnings, more than 30 minutes until people are affected by the tsunami. The incident was probably driven by the proximity of the tsunami source to the city of Palu. This can be seen in large waves that occur in only 3 minutes. A palu-koro fault triggers the proximity of the starting point of the wave as a strike-slip that has a horizontal movement and has a realizing band component in the middle.

Simultaneously, a liquefaction event occurred. This event is quite rare in the world and is very massive, which has made everyone in Indonesia aware of the importance of liquefaction mitigation. It is estimated that the earthquake has caused the collapse of land and buildings on it. The ground turns solid-phase ties, into liquid behaviour and loses its strength. Liquefaction occurs in loose sandy soils and water-saturated mediums which have increased excess pore pressures due to the propagation of earthquake waves to the ground surface.

For this research, liquefaction mapping can be done based on field observation data. The investigative data used is the data from 14,040 boreholes obtained from the Integrated DB Centre of the National Geotechnical Information of the Korea Institute of Construction Technology. For DB data, coordinates and standard penetration test results are used (<http://www.geoinfo.or.kr>). In areas with moderate seismic potential, it can be called rational to make estimations of liquefaction potential zones mapping using standard penetration test results and assessment methods for simplified liquefaction potentials.

1. Analysis of the potential liquefaction

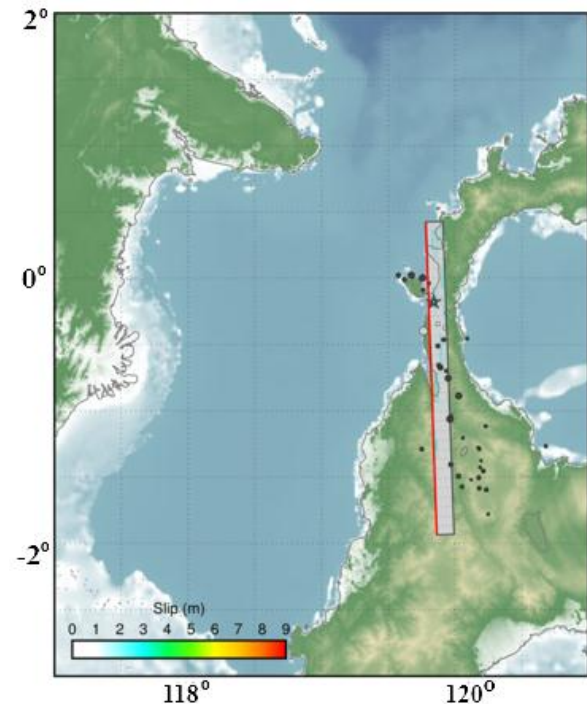


Figure 5. Slip distribution (USGS, 2019)

The results of import data from USGS show that the cross-section of slip distribution of the largest earthquake event on September 28, 2018, extends from the epicentre to the south, as can be seen in Figure 5.

If earthquake data is expanded with a custom magnitude of $M > 5$ from 1900 to 2018, it will have at least 4 earthquake data indicating that the city of Palu may experience liquefaction. Maximum earthquake data in 2018 is sufficient to ensure that potential liquefaction can occur in medium amplitudes only. The data can be seen in Figure 6.

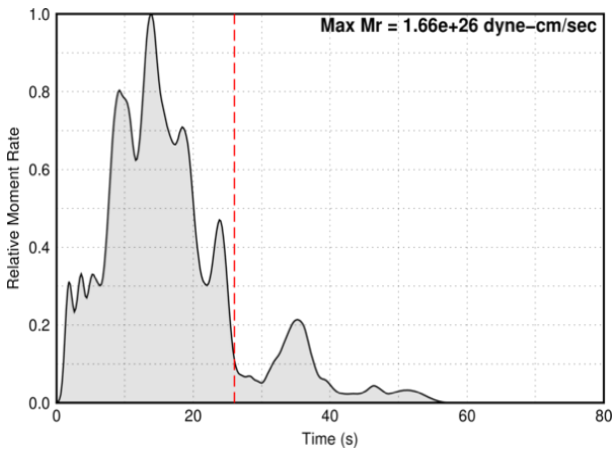


Figure 6. Relative moment rate (USGS, 2019)

2. Mapping and geological criteria

Palu's geological condition is almost entirely in the form of young alluvium, which has the potential to experience liquefaction. The picture of potential conditions due to geological conditions that are correlated with slip distribution shows that the Palu valley area, which is exactly in the Palu-Koro fault area, should get more serious attention.

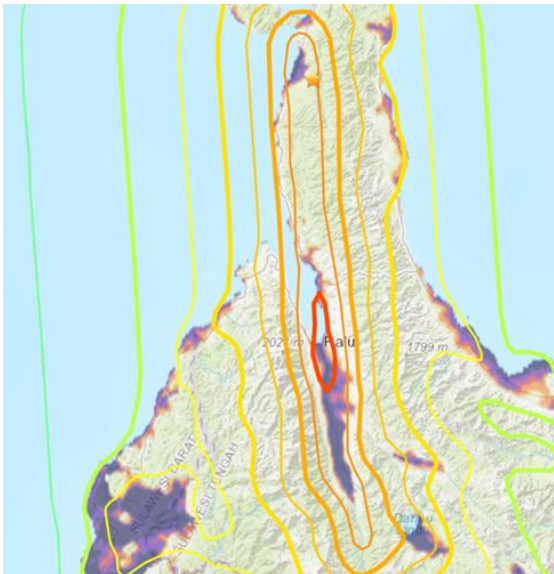


Figure 7. Liquefaction potential map

Mapping potential, slip distribution, criteria, and relative moment rate that have been analyzed show all the liquefaction potential variables can be described properly. The description can be seen in Figure 7.

The final part of this research discussion is measuring all data and mapping that has been compiled with tools owned by a miner. The results obtained are quite significant at the validation stage. The miner library feature is used to describe and get the relative contribution value of each input value. Validation is done by entering earthquake data with a magnitude of $M > 5$ obtained from 1901-2018 from USGS and plotted in the Ambresseys curve. At the validation stage of the prediction, the confirmed model has an R^2 value of 0.89 ± 0.02 , MAD 0.59 ± 0.01 and RMSE 0.53 ± 0.03 with 20 runs performed, while the best hyperparameters to achieve a fit SVM model using $\epsilon = 0.08 \pm 0.02$ and $\gamma = 0.04 \pm 0.01$. At the same time, hyperparameters for ANN are $H = 3 \pm 1$.

5. Conclusions

Utilization of DM for mapping and analysis of big data potential liquefaction through the miner algorithm approach has an excellent capacity to interpret available data. It is also possible to predict the level of risk with several other determinants. The proposed model also makes it possible to identify advanced parameters for the needs of earthquake disaster mitigation in more detail.

REFERENCES

- [1] Al Bawwab, W. M. (2005). Probabilistic assessment of liquefaction-induced lateral ground deformations (Doctoral dissertation, Ph. D. Thesis (C. Özgen. Advisor), Department of Civil Engineering, Middle East Technical University).
- [2] Cortez, P. (2010). Data mining with neural networks and support vector machines using the r/miner tool. *Advances in Data Mining: Applications and Theoretical Aspects*, 10th Industrial Conference on Data Mining, 83. Berlin, Germany: J In P. Perner, editor.
- [3] Faris, A. T., Seed, R. B., Kayen, R. E., & Wu, J. (2006). A semi-empirical model for the estimation of maximum horizontal displacement due to liquefaction-induced lateral spreading. In *Proc. 8th US Conference on Earthquake Engineering* (Vol. 3, pp. 1584-1593).
- [4] Franke, K. W., Ulmer, K. J., Ekstrom, L. T., & Meneses, J. F. (2016). Clarifying the differences between traditional liquefaction hazard maps and probabilistic liquefaction reference parameter maps. *Soil Dynamics and Earthquake Engineering*, 90, 240-249.
- [5] Freitas, A. A. (2013). *Data mining and knowledge discovery with evolutionary algorithms*. Springer Science & Business Media.

- [6] Galavi, V., Petalas, A., & Brinkgreve, R. B. J. (2013). Finite element modeling of seismic liquefaction in soils. *Geotechnical Engineering Journal of the SEAGS & AGSSEA*, 44 (3), 2013.
- [7] Goharriz, M., & Marandi, S. M. (2016). An optimized neuro-fuzzy group method of data handling system based on gravitational search algorithm for evaluation of lateral ground displacements.
- [8] Goh, A. T., & Zhang, W. G. (2014). An improvement to the MLR model for predicting liquefaction-induced lateral spread using multivariate adaptive regression splines. *Engineering Geology*, 170, 1-10.
- [9] Hall, R., & Spakman, W. (2015). Mantle structure and tectonic history of SE Asia. *Tectonophysics*, 658, 14-45.
- [10] Iwasaki, T., Arakawa, T., & Tokida, K. I. (1984). Simplified procedures for assessing soil liquefaction during earthquakes. *International Journal of Soil Dynamics and Earthquake Engineering*, 3(1), 49-58.
- [11] Javadi, A. A., Rezanian, M., & Nezhad, M. M. (2006). Evaluation of liquefaction induced lateral displacements using genetic programming. *Computers and Geotechnics*, 33(4-5), 222-233.
- [12] Kalantary, F., MolaAbasi, H., Salahi, M., & Veiskarami, M. (2013). Prediction of liquefaction induced lateral displacements using robust optimization model. *Scientia Iranica*, 20(2), 242-250.
- [13] Mallick, R., Lindsey, E. O., Kuo, Y. T., Zeng, H., Feng, G., Wang, T., & Hill, E. (2018, December). Large Shallow Slip Along the Palu-Koro Fault Associated with Supershear Rupture. In *AGU Fall Meeting Abstracts*.
- [14] Mohammadnejad, T., & Andrade, J. E. (2015). Flow liquefaction instability prediction using finite elements. *Acta Geotechnica*, 10(1), 83-100.
- [15] Rifai, A. I., Hadiwardoyo, S. P., Correia, A. G., Pereira, P., & Cortez, P. (2015). Data Mining Applied for The Prediction of Highway Roughness under Overloaded Traffic. *International Journal of Technology* 5:751-761
- [16] Rifai, A., Hadiwardoyo, S. P., Correia, A. G., & Pereira, P. (2016). Genetic Algorithm Applied for Optimization of Pavement Maintenance under Overload Traffic: Case Study Indonesia National Highway. *Applied Mechanics and Materials* ISSN: 1662-7482, Vol. 845, pp 369-378 Trans Tech Publications, Switzerland
- [17] Serva, L., Vittori, E., Comerci, V., Esposito, E., Guerrieri, L., Michetti, A. M., & Tatevossian, R. E. (2016). Earthquake hazard and the Environmental Seismic Intensity (ESI) scale. *Pure and Applied Geophysics*, 173(5), 1479-1515.
- [18] Shahri, A. A. (2016). Assessment and prediction of liquefaction potential using different artificial neural network models: a case study. *Geotechnical and Geological Engineering*, 34(3), 807-815.
- [19] Socquet, A., Hollingsworth, J., Pathier, E., & Bouchon, M. (2019). Evidence of supershear during the 2018 magnitude 7.5 Palu earthquake from space geodesy. *Nature Geoscience*, 12(3), 192.
- [20] Sonmez, H., & Gokceoglu, C. (2005). A liquefaction severity index suggested for engineering practice. *Environmental Geology*, 48(1), 81-91.
- [21] Sulewska, M. J. (2017). Applying artificial neural networks for the analysis of geotechnical problems. *Computer-Assisted Methods in Engineering and Science*, 18(4), 231-241.
- [22] Thein, P. S., Pramumijoyo, S., Brotopuspito, K. S., Kiyono, J., Wilopo, W., Furukawa, A., & Putra, R. R. (2015). Estimation of S-wave velocity structure for sedimentary layered media using microtremor array measurements in Palu City, Indonesia. *Procedia Environmental Sciences*, 28, 595-605.
- [23] USGS.gov, <https://earthquake.usgs.gov/earthquakes/map>, August 09, 2019, 14.55.30
- [24] Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. *Knowledge and Data Engineering*, IEEE Transactions on, 26(1), 97-107.
- [25] Youd, T. L., & Perkins, D. M. (1987). Mapping of the liquefaction severity index. *Journal of Geotechnical Engineering*, 113(11), 1374-1392.
- [26] Zhang, G., Robertson, P. K., & Brachman, R. W. I. (2004). Estimating liquefaction-induced lateral displacements using the standard penetration test or cone penetration test. *Journal of Geotechnical and Geoenvironmental Engineering*, 130(8), 861-871.