17th World Conference on Earthquake Engineering, 17WCEE
Sendai, Japan - September 13th to 18th 2020

EVALUATION OF LIQUEFACTION OCCURRENCE FOR LARGE AREA: A MACHINE LEARNING APPROACH

K. Kuwabara(1), H. Xiu(2), T. Shinohara(2), M. Matsuoka(3)

(1) Graduate Student, Department of Architecture and Building Engineering, Tokyo Institute of Technology, kuwabara.k.ab@m.titech.ac.jp

(2) Ph.D. Student, Department of Architecture and Building Engineering, Tokyo Institute of Technology, xiu.h.aa@m.titech.ac.jp, shinohara.t.af@m.titech.ac.jp

(3) Professor, Department of Architecture and Building Engineering, Tokyo Institute of Technology, matsuoka.m.ab@m.titech.ac.jp

Abstract

Soil liquefaction is major cause of damage to soil structures, lifelines, building foundation, and other types of infrastructures. The countermeasures of liquefaction against future earthquakes; i.e. Tonankai-Nankai earthquake being the worst scenario in Japan, are important. Two types of methods for predicting liquefaction damage are widely used. The first one is numerical analysis methods. The methods are often carried out using on-site data. The numerical analysis methods can quantitively predict liquefaction, but it is challenging to apply the method to large areas. The second one is empirical method. These methods can be utilized in large areas. However, previous researches did not consider recent major earthquakes to estimate liquefaction using empirical methods. In this study, we use empirical models predicting the occurrence of liquefaction. An empirical method for predicting liquefaction for large areas is proposed; which is utilizing 40 earthquakes including not only earthquakes with liquefaction over 10 locations from the year 745 up to 2008 but also recent the 2011 Tohoku and the 2016 Kumamoto earthquakes. Our created dataset consists of 11 explanatory variables stored in grid cells measuring 7.5 arc-second latitude × 11.25 arc-second longitude (approximately 250 m × 250 m): geomorphologic classification, elevation, slope angle, relative relief, average shear wave velocity up to 30 m depth (Vs30), distance to the river, distance to the coast, earthquake magnitude, seismic intensity, peak ground velocity (PGV), and peak ground acceleration (PGA). Furthermore, our dataset has highly imbalance problem of the classes, because the number of approximately 18,000 cells are liquefaction grid cells while the number of approximately 130 million cells are non-liquefaction grid cells. The liquefaction grid cells are not uniformly presented; some of the data are presented as the sand boiling area (polygon), while others are recorded as points because the sand boiling areas are unclear. To evaluate both types of data in a uniform manner, the polygon data were transformed to point data by extracting the median point. In this paper, in order to estimate liquefaction occurrence, random forest, which is one of the machine learning methods, was selected to solve the binary classification of liquefaction as opposed to non-liquefaction. In addition, we proposed an ensemble method to obtain a more reliable prediction result. Fifty models were created, and the final decision was defined as the average classification probability over the fifty models. In the training data, the liquefaction class was common in each learning model. The same number of non-liquefaction class as the liquefaction class were extracted by under-sampling based on specific conditions. This process, in which each model learns with the equal proportion of liquefaction and non-liquefaction classes, was effective for imbalanced data. The proposed method achieved an overall accuracy of 88.8%, a recall of 92.4% and a precision of 3.4% with imbalanced data. Finally, a parametric study based on ground motion variables (seismic intensity, PGA, PGV) was conducted to create liquefaction classification maps.

Keywords: liquefaction, empirical method, random forest, imbalanced data, ensemble

1. Introduction

Soil liquefaction is a well-known phenomenon in which soil loses its strength as a solid and flows like a liquid. It is important to take countermeasures against soil liquefaction that disrupts lifeline facility, building foundation and other types of infrastructure. Recent earthquakes such as the 2011 Tohoku earthquake and the 2016 Kumamoto earthquake caused serious damage [1, 2], and there are also concerns about significant damage probability due to future earthquakes such as Tonankai-Nankai earthquake. It is obviously imperative for establishing a robust country to predict future liquefaction. There are two types of methods for evaluating liquefaction damage: numerical analysis methods and empirical methods. Numerical analysis methods are often carried out using on-site data, i.e., logging data, N-value, fine fraction content, and groundwater level along with ground shaking data [3,4]. Therefore, numerical analysis methods are known to quantitively evaluate liquefaction potential at investigated points. However, it is costly to investigate a lot of points. Since it is challenging to adopt numerical analysis methods for large areas, the empirical method is utilized to evaluate liquefaction for large areas.

Especially in Japan, it is typical for an empirical method which evaluates liquefaction to utilize geomorphologic classification, a technique used for zoning the subsurface ground conditions based on a homogeneous landform (relief), formation processes, age, and constituent materials. Wakamatsu (1991) proposed the relationship between geomorphologic classification and liquefaction [5]. Generally, liquefaction is known to occur in terrains where the groundwater level is shallow, and the sand is loosely deposited [6]. In geomorphologic classification, the terrain is mainly comprised of abandoned river channels, filled land, sand dune and lowland between coastal dune and/or bars. In fact, liquefaction occurred at these terrains during past earthquakes [7,8,9]. Matsuoka et al. (2015) proposed a method to evaluate liquefaction potential of each grid cell, which is approximately 250 m × 250 m, over large areas using geomorphologic classification and seismic intensity [10]. However, their work merely relied on 10 earthquakes without including recent events such as the 2011 Tohoku and the 2016 Kumamoto earthquakes. Recent research by Zhu et al. (2017) developed a logistic regression model for predicting liquefaction probability for grid cells, which are approximately 900 m × 900 m, using various parameters, i.e., peak ground velocity (PGV), peak ground acceleration (PGA), average shear wave velocity up to 30 m depth (Vs30), distance to the coast, distance to the river, distance to a water body, water table depth, compound topographic index, precipitation and elevation above the nearest water [11]. Their regression model achieved high performance, but their spatial resolution is coarse and they did not consider geomorphologic classification which is known to be an extensively important factor against liquefaction. To overcome those challenges and improve the performance, it is necessary to predict liquefaction with geomorphologic classification along with other parameters regarding liquefaction. Furthermore, recent earthquakes are considered in this study for more reliable predictions.

Concretely, the objective of this study is to propose a prediction model based on empirical datasets for liquefaction in Japan, using not only geomorphologic classification but also other parameters regarding terrain, soil condition, and earthquake ground motion. In the case of using an empirical dataset, machine learning is well known to be utilized recently in various fields. As a machine learning method, random forest is widely used. Thus, we adopted random forest to solve the binary classification of liquefaction or non-liquefaction.

In this paper, firstly a dataset is shown that includes parameters regarding liquefaction stored in grid cells which are approximately 250 m × 250 m. Then we show processes of making our learning dataset and input ground motion for parametric study. Subsequently, our model is evaluated quantitatively using the confusion matrix, accuracy, precision, recall, f-value. Finally, a parametric study is performed considering earthquake ground motion using our trained model to evaluate future liquefaction potential and to generate liquefaction classification maps.

2. Dataset

2.1 Liquefaction grid data

© The 17th World Conference on Earthquake Engineering

In this research, numerical values in the dataset are stored in grid cell measuring 7.5 arc- second latitude × 11.25 arc-second longitude (approximately 250 m × 250 m). For the data on liquefaction sites, historical data sets for liquefaction induced by earthquakes during the period of 745 up to 2008 [12], the 2011 Tohoku earthquake [1] and the 2016 Kumamoto earthquake [2] were used. For earthquakes that occurred before 2008 [10], 38 earthquakes with liquefaction over 10 locations were included. Among the liquefaction sites, data originating from local artificial changes, such as backfilled manholes and liquefaction of backfills in gravel pits during the 2004 Niigata-ken-chuetsu earthquake [7], were excluded. The liquefaction data are not uniformly presented; some of the data are presented as the sand boiling area (polygon), while others are recorded as points because the sand boiling areas are unclear. To evaluate both types of data in a uniform manner, the polygon data were transformed to point data by extracting the median point. Then, liquefaction grid cells of 250 m which include one or more liquefaction points were defined. Table 1 shows the earthquakes used in this study, while Fig. 1 shows the liquefaction site.

2.2 Explanatory variables

Features in the dataset consist of terrain and earthquake ground motions variables. First, terrain variables include geomorphologic classification, elevation, relative relief, slope angle, Vs30, distance to the river and distance to the coast. The geomorphologic classification and Vs30 data developed by Matsuoka et al. (2015, 2006) were utilized [10,13], and other variables were available at the Ministry of Land, Infrastructure, Transport and Tourism [14]. Relative relief was calculated as the difference between the maximum and minimum elevation across the grid cell of 250 m. Distance to the river and to the coast were calculated as the shortest distance between line and the center of each grid cell. The total number of terrain grid cells is 6.9 million. Earthquake ground motion variables consist of JMA-scale seismic intensity, PGV, PGA, and earthquake magnitude. While magnitude is constant in each earthquake, other variables regarding ground motion were calculated in three different methods. Common features of methods are to calculate seismic ground motions on a reference ground (Vs30 equal to approx. 600 m / s), and then calculate surface ground motions by multiplying an amplification factor based on Vs30 [15]. In other words, there are difference in calculation of the seismic ground motions on the reference ground. The features of three methods are described below.

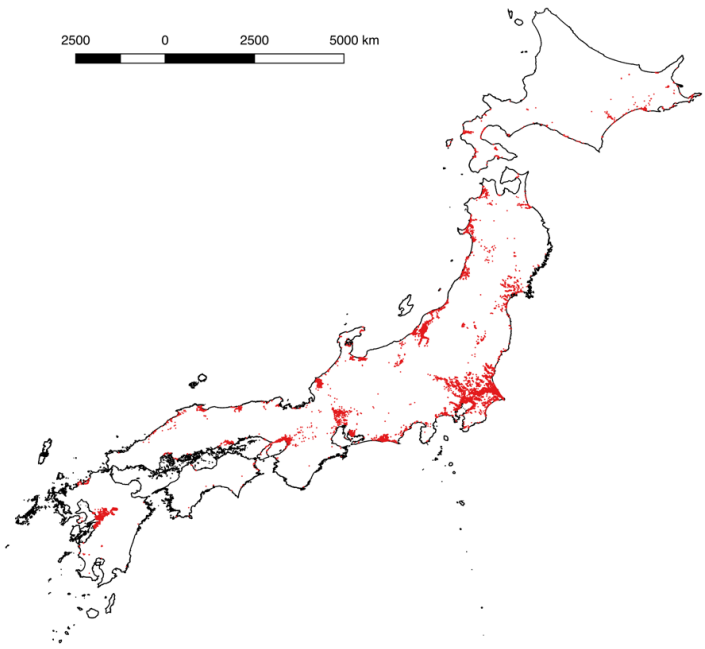


Fig. 1 – Liquefaction distribution in the dataset. The red color indicates liquefaction grid cell.

© The 17th World Conference on Earthquake Engineering

- *Method*-I: The seismic ground motions on the reference ground were calculated using an attenuation relation [16] from the epicenters.
- *Method*-II: The seismic ground motions on the reference ground were calculated using an attenuation relation [16] considering the earthquake faults.

Table 1 - Summary of earthquakes in the liquefaction database

Date	Earthquake	M_w	Number of liquefaction points	Number of liquefaction grid cells	Calculation method
1891 Oct 18	Noubi	8.0	227	226	II
1894 Oct 22	Syonai	7.0	50	50	I
1896 Aug 31	Rikuu	7.2	48	48	I
1897 Jan 17	No-name	5.2	16	16	I
1909 Aug 14	Anegawa	6.8	120	120	I
1923 Sep 1	Kanto	7.9	852	753	II
1927 Mar 7	Kita-Tango	7.3	19	17	II
1931 Sep 21	Nishi-Saitama	6.9	127	124	II
1936 Feb 21	Kawachiyamato	6.4	14	14	I
1939 Sep 21	Oga	6.8	14	14	II
1943 Mar 4	No-name	6.2	12	12	I
1943 Sep 10	Tottori	7.2	90	88	II
1944 Dec 7	Tonankai	7.9	480	470	II
1945 Jan 13	Mikawa	6.8	156	150	II
1946 Dec 21	Nankai	8.0	43	43	I
1948 Jun 28	Fukui	7.1	170	169	II
1952 Mar 4	Tokachi-oki	8.2	15	15	II
1964 May 7	No-name	6.9	18	18	II
1964 Jun 16	Niigata	7.5	268	265	II
1968 May 16	Tokachi-oki	7.9	78	76	II
1978 Jun 12	Miyagi-ken-oki	7.4	52	52	II
1982 Mar 21	No-name	7.1	21	21	II
1983 May 26	Nihonkai-chubu	7.7	345	342	II
1987 Dec 17	Chiba-ken-toho	6.7	335	312	III
1993 Jan 15	Kushiro-oki	7.5	303	140	II
1993 Feb 7	No-name	6.6	76	30	II
1993 Jul 12	Hokkaido-nansei-oki	7.8	506	341	II
1994 Oct 4	Hokkaido-toho-oki	8.2	196	146	II
1994 Dec 28	Sanriku-haruka-oki	7.6	89	59	II
1995 Jan 17	Kobe	7.3	8,083	1,269	III
2000 Oct 6	Tottori-ken-seibu	7.3	418	216	III
2003 May 26	Miyagi-ken-oki	7.1	27	22	III
2003 Jul 26	Miyagi-ken-hokubu	6.4	31	28	III
2003 Sep 26	Tokachi-oki	8.0	148	121	III
2004 Oct 23	Niigata-ken-chuetsu	6.8	1,835	1,395	III
2005 Mar 20	Fukuoka-ken-seiho-oki	7.0	88	73	III
2007 Mar 25	Noto-hanto	6.9	43	35	III
2007 Jul 16	Niigata-ken-chuetsu-oki	6.8	201	112	III
2011 Mar 11	Tohoku	8.4	10,035	8,970	III
2016 Apr 14	Kumamoto	7.3	1,910	1,910	III
	Total		27,406	18,035	

© The 17th World Conference on Earthquake Engineering

- *Method*-III: The seismic ground motions on the reference ground were calculated by interpolation by simple kriging and attenuation relation using observed ground motion data with elimination of the amplification factor [15].

Table 1 also shows the method in each earthquake. The *Method*-III follows the calculation procedure on QuiQuake, National Institute of Advanced Industrial Science and Technology [17]. Finally, the total number of ground motion grid cells is 113 million.

We show processes for making dataset for training and validation using explanatory variables. Since the number of liquefaction class is 18,000 out of 113 million grid cells, it is necessary to reduce the data of the non-liquefaction class to balance the data for learning purpose. Some geomorphologic classifications in which no liquefaction has occurred in any of the 40 earthquakes have been excluded because it was assumed that liquefaction might not occur in these terrain classifications in the future. Fig. 2 shows a boxplot of the relationship between terrain variables and liquefaction, excluding geomorphologic classification. From Fig. 2, these variables are negatively correlated with liquefaction. Therefore, upper limits were set. In each earthquake, these upper limits were defined as maximum values in liquefaction class. The terrain that might cause liquefaction was extracted using the upper limits. As for the ground motions, there is a positive correlation among PGV, PGA and seismic intensity, and thus an effort was made to reduce the data based on seismic intensity. Specifically, the lower limit was set because the seismic intensity has a positive correlation with liquefaction. The lower limit of seismic intensity was defined as 0.5 less than the minimum seismic intensity

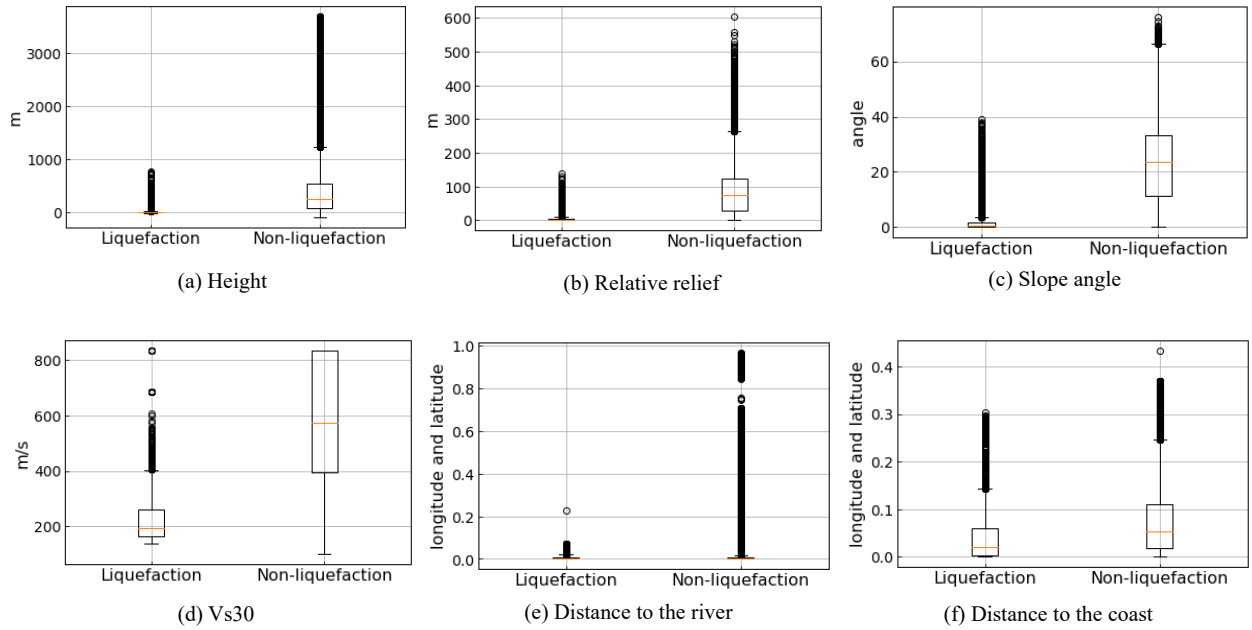


Fig. 2 – Box plots of terrain variables excluding geomorphologic classification in liquefaction and non-liquefaction. The boxes are the quartiles, representing the 25%, 50%, and 75% points from the bottom. The minimum and maximum values are within 1.5 times the length of the box; anything beyond this particular value is plotted as an outlier with black circles.

in which liquefaction occurred for each earthquake. The seismic intensity that could cause liquefaction was extracted using the lower limit. As a result, the terrain grid cells were reduced to about 4.3 million, and the ground motion grid cells were reduced to about 6.9 million.

3. Proposed Methods

3.1 Ensemble learning method

© The 17th World Conference on Earthquake Engineering

Random forest [18], which is one of the most representative machine learning methods, was utilized to predict liquefaction in this study. To mitigate our imbalanced problem, 50 random forest models are trained from different datasets. Each dataset is sampled with replacement from the original one. At the time of prediction, the final result is obtained by taking the arithmetic mean of the prediction results of all models. This method is defined as an ensemble learning in this study.

The sampling method for reducing imbalance problem is presented. As for sampling training data, the data of liquefaction class are the same for all 50 models. The training data is 80% of the liquefaction class and 20% consists of testing data. Because imbalanced data condition has an adverse effect on learning, each model learns at the same class ratio. Therefore, in each model, the number of non-liquefaction class is approximately 14,500 having the same number as the liquefaction class. The proportion of randomly sampled non-liquefaction class and liquefaction class from a geomorphologic classification category is identical because previous study [5] has pointed out that the geomorphologic classification is significantly related to liquefaction. In each earthquake, the described random sampling was performed. Therefore, the non-liquefaction class sampling was carried out from earthquakes where more liquefaction occurred. The non-liquefaction class in test data was randomly sampled at 10% from each geomorphologic classification for each earthquake. Finally, the model accuracy was verified in a liquefaction class of approximately 3,500 and a non-liquefaction class of 800,000.

3.2 Parametric study to evaluate liquefaction occurrence

A parametric study of variables related to ground motion is carried out to estimate future liquefaction occurrence using learned models. In this parametric study, seismic intensity, PGV, and PGA on stiff (reference) ground are changed. The magnitude is defined as the mean of 40 earthquakes. Since the three variables, seismic intensity, PGA, and PGV, are correlated with each other, a non-linear regression analysis was carried out between seismic intensity and PGV or PGV using the exponential function of Eq. (1).

$$y = a\exp(bI) \tag{1}$$

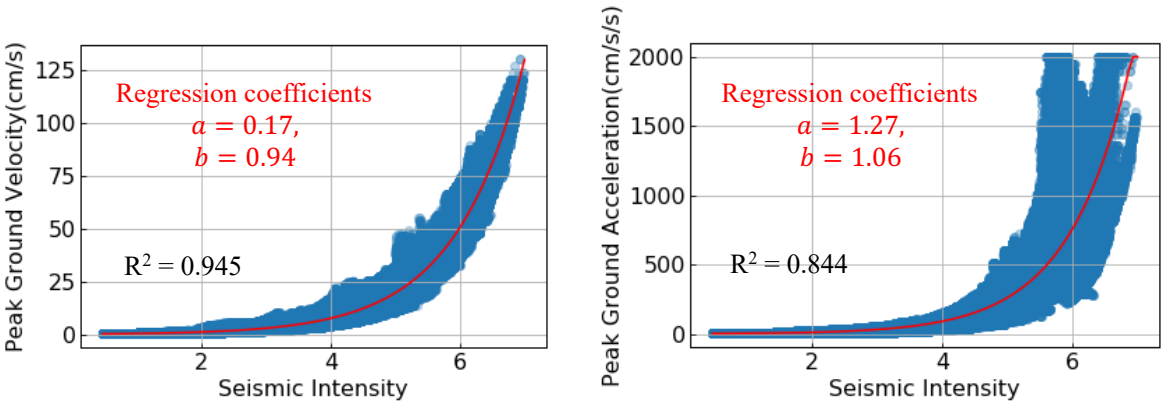
Thus, pseudo seismic ground motions were created to execute the parametric study. I is the seismic intensity. a and b are regression coefficients. The regression analysis was performed on the data of the 2011 Tohoku earthquake because similar result was obtained by regression analysis in all data. In the regression analysis between seismic intensity and PGA, the upper limit was set to 2,000 cm/s/s which is the upper limit of PGA defined by Matsuoka et al. (2015) [10]. Fig. 3 shows the results of regression analysis. The correlation coefficients for determining the seismic intensity and PGV and PGA were 0.95 and 0.84, respectively. In order to simulate more realistic ground motions, this study considers amplification of ground motion based on Vs30. The I_{ref}, which is the seismic intensity on the reference ground, is fluctuated in the parametric study. The amplification is defined as ΔI using an equation for calculating amplification and attenuation based on Vs30 [15]. Thus, the fluctuating ground motion variables are obtained as follows.

$$I = I_{ref} + \Delta I \tag{2}$$

$$PGV = 0.17\exp(0.94\,I) \tag{3}$$

$$PGA = 1.27\exp(1.06\,I) \tag{4}$$

© The 17th World Conference on Earthquake Engineering



(a) Seismic intensity and PGV (b) Seismic intensity and PGA

Fig.3 – The results of regression analysis.

4. Results

4.1 Evaluation of models

A performance of the proposed method was evaluated. The results of the binary classification were evaluated by a confusion matrix. The confusion matrix is composed of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). A threshold for liquefaction was set as 50 % that is liquefaction class probability calculated by taking the arithmetic mean of the prediction results of all models. From these four factors, four performance measurements which are *Accuracy*, *Precision*, *Recall* and *F value* are defined and then calculated according to Eq. (5), Eq. (6), Eq. (7) and, Eq. (8), respectively.

$$Accuracy = \frac{\mathrm{TP} + \mathrm{FP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{5}$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{6}$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{7}$$

$$F\ value = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{8}$$

Table 2 shows the results of the confusion matrix and four performance measurements by the proposed ensemble method and the results obtained by one model. According to Table 2, in the case of the non-ensemble method, the *Recall* was only 0.376 because the learning largely depended on the majority non-liquefaction class meaning that it was not enough to detect liquefaction. On the other hand, our proposed ensemble method shown the *Recall* was 0.924. The detection ability of liquefaction in the proposed method was higher than the non-ensemble method. However, the *Precision* in the proposed method was 0.034. In other words, there were many misclassifications of non-liquefaction classes. In case of disasters such as liquefaction, it is more important to emphasize *Recall* than *Precision*. In order to discuss the validity of detection performance, Table 3 shows the cancer screening investigation in Japan [19] to compare with our result. Our objective is similar to the cancer investigation because both are striving to catch the potentiality generally. In other words, our study and the cancer screening investigation are in the first stage for detection, and have detail detections in the next stage. Compared to our study's *Precision* in Table 3, our *Precision* is higher than that in cancer screening excluding breast cancer. Also, comparing with the model proposed by Zhu et al. (2017) [11], our results were considerately lower than that. The reasons of this problem are inferred that the resolution is 900

© The 17th World Conference on Earthquake Engineering

Table 2 – Confusion matrix and performance index in non-ensemble and ensemble method.

		Non-Ensemble (1 model)		Ensemble (50 models)	
		Actual		Actual	
		Liquefaction	Non-Liquefaction	Liquefaction	Non-Liquefaction
Prediction	Liquefaction	1,345 (TP)	1,275 (FP)	3304 (TP)	92,674 (FP)
	Non-Liquefaction	2,229 (FN)	823,844 (TN)	270 (FN)	732,445 (TN)
Accuracy		0.996		0.888	
Precision		0.513		0.034	
Recall		0.376		0.924	
F-value		0.434		0.066	

Table 3 – Cancer screening investigation in Japan in 2016 year [19].

	Stomach	Lung	Colorectal	Uterine	Breast
Cancer examinee	2,482,333	4,075,104	4,636,731	3,804,714	2,584,439
Required Detailed Inspection (RDI)	168,218	62,193	286,815	80,882	176,836
Cancer Diagnosis (CD)	2,523	1,381	7,943	1,355	7,336
Precision (CD/RDI)	0.015	0.022	0.028	0.017	0.042

m square, and they did not deal with so high imbalanced problem for non-liquefaction class. In other words, their model was not adequate as for validation of test data because their model depended on a fewer non-liquefaction classes than us. Therefore, it cannot be said sweepingly that our model is definitely more inferior.

4.2 Evaluation of liquefaction occurrence

The parametric study of variables related to ground motion, which is shown in 3.2 section to determine the liquefaction occurrence in the target area, was performed using learned models. Specifically, the surface ground motions, *I*, *PGA*, and *PGV*, were created using I_{ref} by Eq. (2), Eq. (3) and Eq. (4), where I_{ref} was changed to 3, 4, 4.75, 5.25, 5.75, 6.25. Then, proposed models calculated the liquefaction class probability using the created surface ground motions. After that, the final class probability was obtained by taking the arithmetic mean of the class probability of all models. It is important to note that the class probability is not equal to an actual liquefaction probability. Fig. 4 and Fig. 5 show the liquefaction confidence and classification maps in each I_{ref}, respectively. Fig. 5 was generated based on the class probability using 50% threshold. Comparing with Fig. 1, which is the distribution of liquefaction considered in this study, and Fig. 4 (c), the area where the liquefaction probability was turning to high was the almost same as the liquefaction grid cells in Fig. 1. However, we could observe the different tendency in Hokkaido region (shown in a red rectangle in Fig. 4 (c) and Fig. 5 (c)). We have not utilized the Hokkaido region for training data as shown in Fig. 1. Nevertheless, our model could predict the part of unseen region, where liquefaction actually occurred in 2018 Hokkaido-iburi-tobu earthquake [20]. Therefore, these results indicate that the proposed models could generally capture the tendency for liquefaction. However, it is important to note that the high-risk regions may not be of high accuracy because the proposed method could not achieve high *Precision*.

© The 17th World Conference on Earthquake Engineering

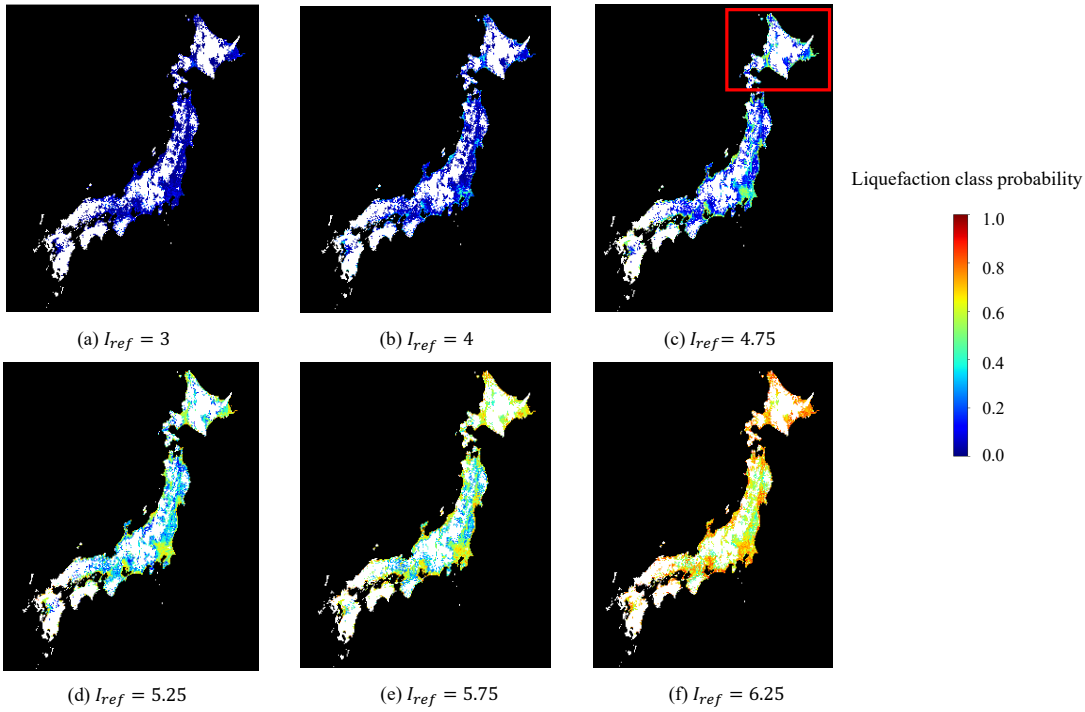


Fig. 4 – Liquefaction confidence map in each I_{ref}. The color and white areas indicate the average liquefaction class probability of all models and the excluded areas, respectively.

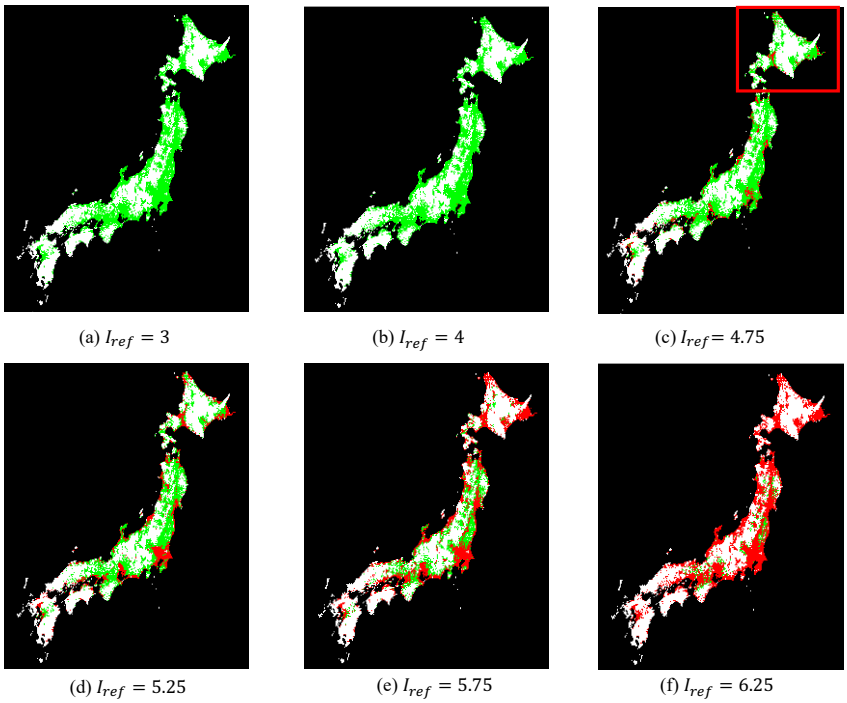


Fig. 5 – Liquefaction classification map in each I_{ref}. The green, red and white areas indicate the non-liquefaction, liquefaction classified by our models and the excluded areas, respectively.

© The 17th World Conference on Earthquake Engineering

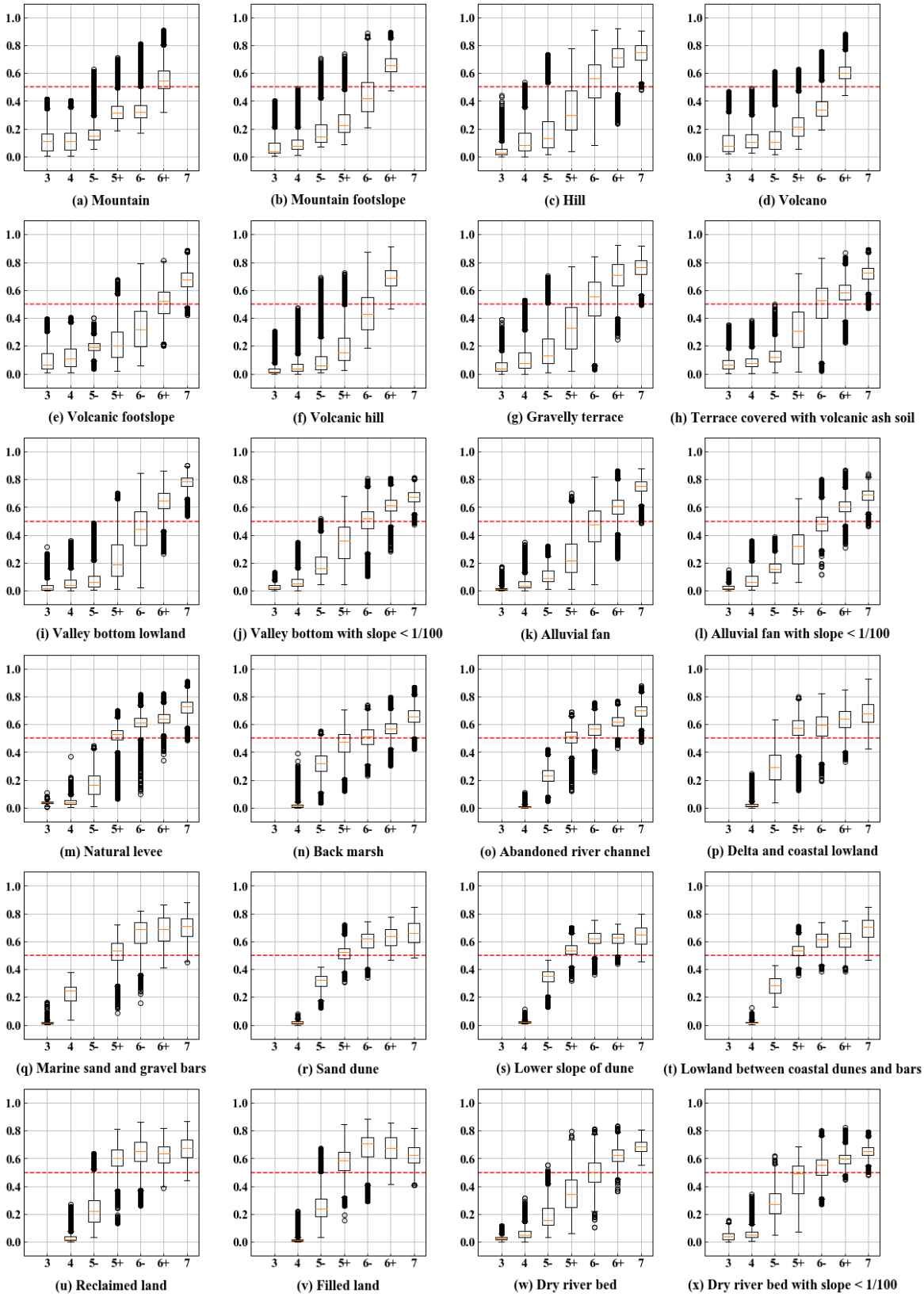


Fig. 6 – Liquefaction class probability in each geomorphologic classification. In each figure, x axis is *I*, and y axis is liquefaction class probability. The red dotted line indicates the threshold for liquefaction.

© The 17th World Conference on Earthquake Engineering

Fig. 6 shows the box plots of liquefaction class probability in each geomorphologic classification. Although there are variations due to other variables, the degree of liquefaction hazard can be evaluated in each geomorphologic classification from Fig. 6. According to that figure, in some geomorphologic classifications, medians of liquefaction class probabilities were over 50% when the *I* was 5+. The geomorphologic classifications are (m) Natural levee, (o) Abandoned river channel, (p) Delta and coastal lowland, (q) Marine sand and gravel bars, (r) Sand dune, (s) Lower slope of dune (t) Lowland between coastal dunes and bars, (u) Reclaimed land, (v) Filled land and (x) Dry river bed with slope < 1/100. A common feature of these classifications is that the groundwater level is shallow. Also, comparing Fig. 6 (i) and (j), (k) and (l), and (w) and (x), the lower land had higher liquefaction class probability as same as the real tendency. When the *I* was over 6+, the median was over 50% in all geomorphologic classifications. In addition, there were also a lot of liquefaction class in Fig. 5 (f). However, it is difficult to consider that liquefaction occur in the geomorphologic classifications such as (a) Mountain and (c) Hill in reality even if it would be with strong ground motion such as the 6+ of seismic intensity. It can be inferred that liquefaction might not occur in the classifications in fact because the classifications were merely representative of the grid cell. In addition, the strong ground motions have rarely occurred in the classifications in the past. Therefore, it can be said that the model generated in this study is only based on the learning data.

5. Conclusion

This study evaluated liquefaction occurrence using 40 earthquakes which occurred from 1891 to 2016 year with liquefaction over 10 locations in Japan. Specifically, the binary classification of liquefaction or non-liquefaction was solved by training the random forest model using 11 variables, which were geomorphologic classification, elevation, slope angle, relative relief, Vs30, the distance to the river, the distance to the coast, earthquake magnitude, seismic intensity, PGV, and PGA. Furthermore, the ensemble learning method was proposed using 50 random forest models in order to tackle the imbalance problem. As a model performance evaluation, *Precision* and *Recall* were 3.4% and 92.4%, respectively. It can be stated that many non-liquefactions are classified as liquefaction although liquefaction is rarely overlooked. In order to discuss our results, we compared with cancer screening investigation in Japan because it was similar to our objective detecting in the first stage. The *Precision* was higher than almost the cancer screening's *Precision*. Finally, liquefaction classification maps were generated using the learned model. As results of parametric study, the region having the risk of liquefaction was almost the same as liquefaction history from 1891 to 2016 year excluding Hokkaido region. However, in 2018 Hokkaido-iburi-tobu earthquake, a lot of liquefaction was actually confirmed in a part of suspected Hokkaido region. For this reason, the trained models ware able to generally capture the tendency of liquefaction whereas it may not be true that our models are completely reliable because of low *Precision*. In the future, this problem is going to be addressed.

6. Acknowledgements

I would like to express my gratitude to all relevant parties. The terrain data excluding geomorphologic classification and Vs30 were provided by Ministry of Land, Infrastructure, Transport and Tourism. The liquefaction survey data of the 2011 Tohoku and the 2016 Kumamoto earthquakes were provided by Dr. Kazue Wakamatsu and Dr. Shigeki Senna.

7. References

[1] Wakamatsu K, Senna S, Ozawa K (2017): Liquefaction and its Characteristics during the 2011 Tohoku Earthquake, *Journal of Japan Association for Earthquake Engineering*, **17** (1), 43-62. (in Japanese)

[2] Wakamatsu K, Senna S, Ozawa K (2017): Liquefaction and its Characteristics during the 2016 Kumamoto Earthquake, *Journal of Japan Association for Earthquake Engineering*, **17** (4), 81-100. (in Japanese)

[3] Architectural Institute of Japan (2001): *Building foundation structure design guidelines*. (in Japanese)

© The 17th World Conference on Earthquake Engineering

[4] Iwasaki T, Tokida K, Tatsuoka F (1981): Soil liquefaction potential evaluation with use of the simplified procedure, *International Conference on Recent Advances in Geotechnical Earthquake Engineering and Soil Dynamics*, St. Louis, 209-214.

[5] Wakamatsu K (1991): Geomorphological evaluation of the subsurface geotechnical ground conditions affecting the occurrence of liquefaction, *Journal of Structural and Construction Engineering,* Architectural Institute of Japan, **421**, 29-37. (in Japanese)

[6] Tohno I (1992): Performance of disaster of soil liquefaction research, *Journal of Japan Society of Agriculture*, **60** (9), 7-12. (in Japanese)

[7] Wakamatsu K, Yoshida N, Kiku H, (2006). Liquefaction during the 2004 Niigata-ken Chuetsu earthquake: General aspect and geotechnical and geomorphological conditions, *Journal of Japan Society of Civil Engineers* **62**, 263–276. (in Japanese)

[8] Senna S, Hasegawa N, Maeda T, Fujiwara H (2012): Liquefaction of the Tonegawa basin caused by the 2011 of the pacific cast of tohoku earthquake, *Journal of Japan Association for Earthquake Engineering*, **12** (5), 143-162. (in Japanese)

[9] Aoyama M, Koyama T, Une H (2014): Geomorphological Condition and Land History of Liquefaction Damaged Sites in the Lower Part of the Tone River Lowland Induced by the 2011 off the Pacific Coast of Tohoku Earthquake, *Geographical Review of Japan Series,* **87** (2), 128–142. (in Japanese)

[10] Matsuoka M, Wakamatsu K, Hashimoto M, Senna S, Midorikawa S (2015). Evaluation of liquefaction potential for large areas based on geomorphologic classification, *Earthquake Spectra*, **31**, 2375–2395.

[11] Zhu J, Baise L. G, Thompson E. M (2017): An updated geospatial liquefaction model for global application, *Bulletin of the Seismological Society of America*, **107** (3), 1365–1385; doi:10.1785/0120160198.

[12] Wakamatsu K (2011): *Maps for Historic Liquefaction Sites in Japan 745-2008 (with DVD- ROM)*, The University of Tokyo Press, 90. (in Japanese)

[13] Matsuoka M, Wakamatsu K, Fujimoto K, Midorikawa S (2006): Average shear-wave velocity mapping using japan engineering geomorphologic classification map, *Journal of Structural Engineering and Earthquake Engineering*, Japan Society of Civil Engineers **23** (1), 57s–68s.

[14] Ministry of Land, Infrastructure, Transport and Tourism: National Land Numerical Information download service, http://nlftp.mlit.go.jp/ksj/

[15] Midorikawa S, Komazawa M, Miura H (2008): Relationship between average shear-wave velocity and site amplification factors obtained by records from Yokohama dense strong-motion network, *Journal of Japan Association for Earthquake Engineering*, **8** (3), 19-30. (in Japanese)

[16] Si H, Midorikawa S (1999): New attenuation relationship for peak ground acceleration and velocity considering effects of fault type and site condition, *Journal of Structural and Construction Engineering*, **523**, 63-70. (in Japanese)

[17] National Institute of Advanced Industrial Science and Technology: Quick Estimation System for Earthquake Maps Triggered by Observation Records, https://gbank.gsj.jp/QuiQuake/QuakeMap/index.en.html.

[18] Habe H (2012): Random Forest, *Information Processing Society of Japan Research Report*, **182** (31).

[19] Ministry of Health, Labor and Welfare (2019): Overview of 2017 Community Health and Health Promotion Business Report, 15. (in Japanese)

[20] Mainichi Newspaper Article (2018): About the liquefaction point occurred by 2018 Hokkaido iburi tobu earthquake, https://mainichi.jp/articles/20181205/k00/00m/040/311000c

© The 17th World Conference on Earthquake Engineering