IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 29, 2021, accepted February 8, 2021, date of publication February 18, 2021, date of current version February 26, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3060114

Predictive Analysis of the Building Damage from the 2011 Great East Japan Tsunami Using Decision Tree Classification Related Algorithms

KUMPOL SAENGTABTIM[1], NATT LEELAWAT[1,2], (Senior Member, IEEE), JING TANG[2,3], WANIT TREERANURAT[4], NARUNPORN WISITTIWONG[4], ANAWAT SUPPASRI[5], KWANCHAI PAKOKSUNG[5], FUMIHIKO IMAMURA[5], NORIYUKI TAKAHASHI[6], AND INGRID CHARVET[7]

[1]Department of Industrial Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand
[2]Disaster and Risk Management Information Systems Research Group, Chulalongkorn University, Bangkok 10330, Thailand
[3]Faculty of Engineering, International School of Engineering, Chulalongkorn University, Bangkok 10330, Thailand
[4]School of Management Technology, Sirindhorn International Institute of Technology, Thammasat University, Pathum Thani 12121, Thailand
[5]International Research Institute of Disaster Science, Tohoku University, Sendai 980-8572, Japan
[6]Department of Architecture and Building Science, School of Engineering, Tohoku University, Sendai 980-8579, Japan
[7]Praedicat, Inc., Los Angeles, CA 90016, USA

Corresponding author: Natt Leelawat (natt.l@chula.ac.th)

This work was supported in part by the Ratchadapisek Sompoch Endowment Fund (2020), Chulalongkorn University "Predictive Analysis Model Development for the Building Damage from Tsunami by Decision Tree Algorithm" under Grant CU_GR_63_39_21_07, in part by JSPS Grant-in-Aid for Young Scientists (B) "Applying developed fragility functions for the Global Tsunami Model (GTM)" under Grant 16K16371, and in part by JSPS-NRCT Bilateral Research Grant.

ABSTRACT When considering a tsunami disaster, many researchers have considered the tsunami's flow depth and velocity as the primary contributors to the building damage. Additionally, the majority of these studies have used the maximum value as the measure of each of these two factors. However, building damage may not occur when the maximum flow depth and the maximum flow velocity of the tsunami are reached. This study addressed two objectives based on the 2011 Great East Japan Earthquake and Tsunami. Firstly, to find out whether the maximum values of the flow depth and flow velocity are the same as their critical values and, secondly, to verify which combination of the parameters is the best predictor of the building damage level. The data from 18,000 buildings in Ishinomaki City, Japan, with the cooperation of the Japanese joint survey team, were analyzed using the decision tree related algorithms. The critical variables were the simulated data at the time when the buildings collapsed. The analysis showed the accuracy of the prediction based on the group of variables. Finally, the findings showed that the combination of the critical flow depth and maximum flow velocity provided the highest accuracy for classifying the level of building damage.

INDEX TERMS Building damages, data mining, decision tree algorithm, 2011 Great East Japan earthquake and tsunami.

I. INTRODUCTION

During the 2011 Great East Japan Earthquake (GEJET), a tsunami was triggered by the magnitude-9 earthquake in the northern area of Japan, which created one of the most disastrous situations in Japan. More than 400,000 buildings were damaged or destroyed and most of the people in the impact area lost their homes and belongings. After this event, researchers in the field of disaster prevention and management examined the damage level of buildings. Attempts were made to understand its characteristics and to come up with variables that could predict the damage level to the buildings.

The associate editor coordinating the review of this manuscript and approving it for publication was Shuihua Wang.

The common variables widely used to predict the damage level to buildings are the tsunami's inundation depth and flow velocity along with the structural materials of the building [1]. Subsequently, the tsunami's maximum flow depth (*Dmax* and maximum flow velocity (*Vmax* have been used as alternative variables to predict the damage level to buildings, as these two variables are measured at the time the buildings collapsed. Recently, some researchers have suggested some alternative approaches [2] that take the load-resistance into account. Thus, some recent studies shifted the key variables to the

TABLE 1. Building damage level for earthquakes and tsunamis.

Ministry of Land, Infrastructure, Transport and Tourism (MLIT)[a]	Japan Prime Minister's Office[b]	EMS-983 (European macroseismic scale)[b]	Architectural Institute of Japan[b]	Federal Emergency Management Agency (FEMA [7])[c]
Rank 0: No damage	No damage	Grade 0: No damage	Rank 0: No damage	Slight
Rank 1: Minor	Moderate damage	Grade 1: Slight damage	Rank 1: Negligible damage	Moderate
Rank 2: Moderate	Heavy damage	Grade 2: Moderate damage	Rank 2: Slight damage	Extensive
Rank 3: Major	Major damage	Grade 3: Heavy damage	Rank 3: Moderate damage	Complete
Rank 4: Complete		Grade 4: Very heavy	Rank 4: Major damage	
Rank 5: Collapsed		Grade 5: Destruction	Rank 5: Collapse	
Rank 6: Washed away				

[a]Ministry of Land, Infrastructure and transportation (MLIT) [8]. Survey of tsunami damage condition. Retrieved August 25, 2020, from https://www.mlit.go.jp/toshi/toshi-hukkou-arkaibu.html
[b]Okada, S., & Takai [9]. Classifications of structural types and damage patterns of buildings for earthquake field investigation. In Proceedings of the 12th world conference on earthquake engineering (paper 0705), Auckland.
[c]FEMA [7]. Hazus Tsunami Model Technical Guidance, FEMA.

critical flow depth (*Dc* and the critical flow velocity (*Vc*, which are the simulated measurement of the tsunami flow at the time when the building collapsed [2], [3].

Several studies have explored how to assess the level of building damage from the 2011 GEJET. Suppasri *et al.* [4] conducted a building damage assessment by calculating the fragility estimation based on the hydrostatic and hydrodynamic forces, and so on. The fragility estimation of their study also included the building characteristics, such as the building material and height. Latcharote *et al.* [5] used the overturning moment concept to illustrate the failure mechanism of an overturned building in Onagawa. The main objective of their study was to determine whether the hydrodynamic or buoyant force of the flow velocity and depth of the tsunami could create a higher overturned moment to the buildings. Macabuag [6] studied the fragility function and proposed a way to identify the key tsunami intensity measures, which include the force that was simulated from the velocity and depth, flow regime, and so on.

In this research, the tsunami flow velocity and depth from the 2011 GEJET were selected as the main variables for classifying the building damage level using a new classification approach. The objective of this study was to provide a predictive analysis by using the novelty machine learning called decision tree related algorithms (DTRAs) [decision tree (DT), random forest (RF), and gradient boosted tree (GBT)] instead of statistical technique to classify the level of building damage based on the *Dc* and *Vc* of the tsunami. By using these DTRAs, this study analyzed whether the simulated values of the *Dmax* and *Vmax* could create the same impact to the buildings as the critical valuesof the tsunami (*Dc* and *Vc*) or not. Additionally, this research verified which combination of variables was the best group to predict the damage level of buildings. In this case, the groups of variables consisted of *Dc* and *Vc*, *Dc* and *Vmax*, *Dmax* and *Vc*, and *Dmax* and *Vmax*. Based on this, this study aims to provide some suggestion to the insurance and related organization about the factors that can be used for estimating the building damage levels from the tsunami.

This article is divided into six main sections. Section two provides a literature review of the classification of building damage in different types of disasters, and the impact factors of building damage. Section 3, the research design and methodology, shows how this research used the variables to predict the damage level based on the DTRAs. Section 4 presents the findings from the analysis, while the discussion and conclusion are provided in Sections 5 and 6, respectively. The implementation of the methods in practice and its limitations are mentioned as well.

II. LITERATURE REVIEW

This section provides an overview of the related studies. They include the building damage from other disasters, building damage factors in water-related disasters, and as the methodologies used in the literature: building damage assessment and classification algorithms.

A. CLASSIFICATION OF BUILDING DAMAGE FROM OTHER DISASTERS

Table 1 summarizes the classification of building damage, based on different types of disasters in different countries. Obviously, there is no standard definition of the damage level of buildings, not just across different disaster types, but also across different countries and organizations in charge of data collection.

The building damage level for the 1985 Mexico earthquake was ranked by the Architectural Institute of Japan, with six ranks comprised of ''No damage'', ''Negligible damage'', ''Slight damage'', ''Moderate damage'', ''Major damage'', and ''Collapse''. Okada and Takai [9] studied the 1995 Hyogo-ken Nanbu earthquake and the 1985 Mexico earthquake. The statistics of the building damage scale from the 1995 Hyogo-ken Nanbu earthquake was organized by Japanese Prime Minister's Office and the scale included the four damage levels of ''No damage'', ''Moderate damage'', ''Heavy damage'', and ''Major damage''. This type of damage scale can only be used for measuring the damage level of the wooden framed buildings. From the same study, the European macroseismic scale (EMS)-98 damage scale for earthquakes was also introduced. This type of the damage level was first developed by Grünthal [10]. Leelawat *et al.* [11] analyzed the building damage from the 2004 Indian Ocean Tsunami at Sri Lanka, following the damage levels and classification description of Murao and Nakazato [12], [13] as ''No/slight damage'', ''Moderate damage'', ''Heavy damage'', and ''Complete damage''.

This study was based on the 2011 GEJET, and so we selected the building damage classification of the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) to be used in the data collection phase, following that of many previous publications [1], [2], [11], [14]–[16]

B. BUILDING DAMAGE FACTORS IN WATER-RELATED DISASTERS

As summarized in Table 2, many studies have estimated the level of damage from water-related disasters. The combination of flow velocity and depth have been used for indicating the flood damage of the Elbe catchment flood in Germany [17]. Similarly, Koshimura *et al.* [18] used the combination of inundation depth, current velocity, and hydrodynamic force to predict the damage of structures by a fragility function. Moreover, the distance from the shoreline has also been used as one of the main variables to estimate the damage levels from a tsunami [19]. On the other hand, some studies have tried to find a way for preventing the tsunami damage by considering the density of the forest and diameter of tree trunks [20].

Based on the factors explained in Table 2, most of the previous studies focused on flow depth and flow velocity of the tsunami, and the considered parameters are the maximum value. In addition, there are limited studies related to the use of the combination between difference variables. Moreover,

TABLE 2. Summary of impact factors for water-related disater.

Factor	Definition	Source(s)
Flow depth	Flow depth is a parameter to determine the impact of the disaster event when it is at a low damage state. It can be measured from the water marks on the standing building.	Kreibich *et al.* [17], Koshimura *et al.* [18], Petrone *et al.* [21]
Flow velocity	Flow velocity is the variable that is predicted from the Bernoulli principle applied to the flow depth before and after an obstacle or by combining measures of the flow depth and Froude number.	Kreibich *et al.* [17], Koshimura *et al.* [22], De Risi *et al.* [23]
Building materials	Building materials are the materials that are used to construct the building, such as wood, brick, concrete, steel, and so on. However, all these materials also help to prevent the structure from the unexpected disasters, like a tsunami. The damage level is different for each type of material used to form the building's structure.	Leelawat *et al.* [1], Suppasri *et al.* [14]
Floors	The number of floors is important in investigating the impact of the tsunami, as the damage to a building from a tsunami is different for buildings with a different number of stories.	Leelawat *et al.* [1], Leelawat *et al.* [24], Suppasri *et al.* [14]
Debris impact	Debris impact is the dynamic load consisting of a rapid impulse being applied on a structure by an object with a certain mass. It consumes energy from the impacting debris, reducing the highest impact force.	Chatvet *et al.* [25], Stolle *et al.* [26], FEMA [7]
Soil material	Tsunami has a great impact on soil erosion and causes a scouring action under the foundation of the structure. The foundation of buildings depends on the nature of the sub soil and keeps the buildings balanced against the impact of the tsunami.	Bikçe *et al.* [27]
Distance from the shoreline	The length of the tsunami wave from the shoreline in a horizontal distance has a powerful influence on the damage level estimation. The tsunami wave gets higher and generates a more destructive potential when it reaches the shore, and provides a strong impact that can damage the structure of buildings nearby to the shore, but this decline as it travels inland. When the tsunami wave comes to the shore it is called the run up, which is the maximum vertical height above the sea level.	Iverson and Prasad [19]
Type of construction	The hydrodynamic forces damage individual structural elements as well as the overall structural system depending upon the construction.	FEMA [7]
Soil material	Tsunami has a great impact on soil erosion and causes a scouring action under the foundation of the structure. The foundation of buildings depends on the nature of the sub soil and keeps the buildings balanced against the impact of the tsunami.	Bikçe *et al.* [27]
Distance from the shoreline	The length of the tsunami wave from the shoreline in a horizontal distance has a powerful influence on the damage level estimation. The tsunami wave gets higher and generates a more destructive potential when it reaches the shore, and provides a strong impact that can damage the structure of buildings nearby to the shore, but this decline as it travels inland. When the tsunami wave comes to the shore it is called the run up, which is the maximum vertical height above the sea level.	Iverson and Prasad [19]
Type of construction	The hydrodynamic forces damage individual structural elements as well as the overall structural system depending upon the construction.	FEMA [7]
Soil material	Tsunami has a great impact on soil erosion and causes a scouring action under the foundation of the structure. The foundation of buildings depends on the nature of the sub soil and keeps the buildings balanced against the impact of the tsunami.	Bikçe *et al.* [27]
Distance from the shoreline	The length of the tsunami wave from the shoreline in a horizontal distance has a powerful influence on the damage level estimation. The tsunami wave gets higher and	Iverson and Prasad [19]

TABLE 2. ***(Continued.)*** **Summary of impact factors for water-related disater.**

	generates a more destructive potential when it reaches the shore, and provides a strong impact that can damage the structure of buildings nearby to the shore, but this decline as it travels inland. When the tsunami wave comes to the shore it is called the run up, which is the maximum vertical height above the sea level.	
Type of construction	The hydrodynamic forces damage individual structural elements as well as the overall structural system depending upon the construction.	FEMA [7]

the use of machine learning techniques for estimating building damage level are currently limited in this area. In this research, the critical value together with the maximum value of the depth and velocity of the tsunami are used as the main factors for classifying the building damage level.

C. BUILDING DAMAGE ASSESSMENT

There have been many studies on how a tsunami can damage buildings. Koshimura and Kabaya [28] introduced a new method for assessing the structural damage using the tsunami fragility or fragility function, based on the inundation depth, velocity, and hydrodynamic force [22]. Tsunami fragility is expressed in terms of the probability of the structure to be destroyed by the tsunami. After that, Koshimura *et al.* [28] introduced the inundation height as a new feature for creating the fragility function, based on the 1993 Hokkaido Nansei-oki earthquake tsunami. Similarly, Mas *et al.* [29] used the fragility function to predict the tsunami damage for the 2010 Chilean Tsunami in Dichato. The same concept of the fragility curve was also presented by Suppasri *et al.* [30] using the data from the 2011 GEJET. The fragility curve was created using linear regression for each building's material, the building damage level, and the inundation depth[30].In addition, Suppasri *et al.* [31] also used the fatality ratio together with a field survey in the Miyagi Prefecture, Japan, which was based on previous research, in order to understand the damage characteristics from this tsunami. Thus, most research has focused only on finding the building damage level based on statistical and qualitative analysis.

D. CLASSIFICATION ALGORITHMS

In this research, the characteristic of the damage level was based on the MLIT criteria. Accordingly, seven classes of damage levels were analyzed using the velocity and inundation depth of the tsunami. Currently, there are a lot of classification techniques that can be used, such as logistic regression, neural network, decision tree (DT), etc.

1) DECISION TREE (DT)

In this research, the DT classification algorithm was selected to be the method for classifying the damage level of the 2011 GEJET due to its simplicity and ease of use for non-experts and that it is the most widely used algorithm [32]. It does not require any assumption for performing the analysis [33], and it is said to be one of the most often chosen classification algorithms [34] since it is considered to be easy and efficient to use for the classification task [35]. The main concept of the DT is to create a tree-like structure that is composed of the root, branch, and leaf node. The top part of the DT is called the ''top root node'', which indicates the most important feature for classifying the label of the data. The DT has the objective to obtain the maximum information and retrieve the lowest amount of entropy or the impurity of the tree. The DT will split its branch from the top node of the tree until the bottom leaves nodes of the tree, which are then the classification result based on the DT algorithm.

2) RANDOM FOREST (RF)

The RF algorithm is a classification method that performs a similar process to DT. The concept of RF was first developed by Breiman [36] and can be called as an ensemble. This concept uses the same concept as the DT and then aggregates the results of each decision tree to obtain the result. The RF algorithm also uses an extension approach, called a bagging approach, where the different features of the data set will be assigned to each decision tree. This randomization of the feature can also be called a ''bootstrapping concept'' [37]. Thus, the best result can be obtained [38]. In addition, RF can reduce the ''overfitting problem'' based on the feature selecting criteria that can otherwise create some interdependence [39]. Therefore, the RF can create a good classification result that can compare to the result of the ''Support Vector Machine'' (SVM) [40], [41].

3) GRADIENT BOOSTED TREE (GBT)

The GBT algorithm is a classification method that was first developed by Schapire *et al.* [42]. This concept is called boosting and uses successive learners to learn the mistake from the previous results. The concept of GBT is also another classification DTRA that performs in a similar way to obtain the result as the DT and RF algorithms. However, GBT uses an iterative algorithm that creates a better decision tree by learning the error from the previous decision tree that had already been created [43]. Thus, GBT is a powerful, fast learning algorithm for the classification and regression problems and it has been used for many commercials such as fraud detection [43].

The three DTRAs identified in this section (DT, RF, and GBT) are very useful for analyzing the main variables that impact on the level of building damage based on a machine learning technique.

III. RESEARCH DESIGN AND METHODOLOGY

The research design, including the study area information, data collection, and descriptive statistics, as well as methodology of this research, are explained in this section.

A. STUDY AREA AND DATA COLLECTION

The data on the damaged buildings used in this research were collected in Ishinomaki City, in the Northeastern area of Japan, by the Tohoku Earthquake Tsunami Joint Survey Group [44]. In addition, this research used the MLIT survey and other survey reports as well. Some members of this research also took part in the survey group and provided a lot of the data set. The *Dc*, *Vc*, *Dmax*, and *Vmax* variables were derived by simulation, as explained in detail in [2]. In contrast, the damage levels were recorded during the field survey after the 2011 GEJET, so the damage that would be caused considering the simulated value might be different from the real damage caused by the tsunami flow. The damage level for each building was specified based on the MLIT criteria. The data for damaged buildings in this research were only for first-floor wooden buildings, because in order to simulate the critical flow value of the tsunami, the simulated buildings must be easily destroyed. Based on the MLIT criteria, the damage levels were classified by the structural and non@hyphestructural damage of the building. Therefore, if the structural components of the building are damaged, the damage level of the building will be classified at the "washed away", "collapsed", and "complete damage" levels. However, if only the non@hyphestructural components of the building were damaged, the damage level of the building would be classified as at a "moderate" or "minor" level [30].

B. DESCRIPTIVE STATISTICS

Based on the field survey and simulation analysis by the Japanese research team, the data related to the building damage level were gathered and cleaned before performing the analysis. The data cleansing process was performed by deleting some unrelated data and unrealistic data based on the simulation and field survey. After the data cleaning process, the cleaned data were sent to Thai research team based in an excel spread sheet format. The four main variables used in this research were the *Dc*, *Vc*, *Dmax*, and *Vmax*. The average, standard deviation (SD), maximum, and minimum value for each variable are summarized in Table 3. The average critical variables *Dc* and *Vc*) were lower than the maximum variables *Dmax* and *Vmax*). Since the fracture or breaking point of the material happened after the material had reached its ultimate strength point [45], then the critical value of the tsunami flow was reached after the tsunami flow had reached its maximum value. The maximum value of both *Dc* and *Dmax* were 5.789 and 8.128 m, respectively, because the Japanese survey team only simulated for the first floor of the wooden building. The maximum *Vc* and *Vmax* were 6.611 ms and 23.8 kmh, respectively. This is the characteristic of the tsunami flow because as the tsunami moves up to the land, the velocity will decreasedue to the wave height of the tsunami. Finally, in this research, the no damage level, or buildings that received no damage, were classified as damage level 1. Then, "*Minor*", "*Moderate*", "*Major*", "*Complete*", "*Collapse*", and "*Washed away*" damage levels were classified as damage levels 2, 3, 4, 5, 6, and 7, respectively.

TABLE 3. Descriptive statistic of the four main variables used in this study.

	Dc (M)	*Vc* (m/s)	*Dmax* (m)	*Vmax* (m/s)
Mean	0.651	0.639	1.707	0.943
SD	0.515	0.482	1.487	0.577
Max.	5.789	6.611	8.128	6.611
Min.	0.000	0.000	0.002	0.000

C. METHODOLOGY: CLUSTERING METHOD

This research used *RapidMiner Studio Version 9.0.003* for performing the analysis. Based on the variables in this study, the values of each variable could notbe interpreted using the DT algorithm directly because this type of algorithm does not fit with continuous variables. Therefore, a clustering method was the proper way for grouping the same characteristic data into a meaningful group. Due to its popularity and wide usage, the *K@hyphemean clustering algorithm* was applied in this analysis [46]. This algorithm can separate the data into k different clusters and set the center (called the "centroid point") for each cluster. Then, it groups the data into clusters by selecting the data that have a distance, called the "Euclidian distance", closest to the center for each cluster. The distance between the data and the center of the cluster was minimized using (1) [47];

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^j - c_j \right\|^2 \quad (1)$$

where $x_i^{(j)}$ is the value of the variable x for observation i, and c_j is the center point of cluster j.

Based on the clustering algorithm, only two variables were selected each time for analyzing in the clustering process. The optimal number of clusters can be checked using the "*Davies Bouldin Index" (DBI)* [48]. A lower DBI indicates a better performance of the clustering algorithm based on the specified k clusters. From Table IVa, the *Dc* & *Vc* paired variables were clustered into two groups, while the other paired variables (*Dc* & *Vmax*, *Dmax* & *Vc*, and *Dmax* & *Vmax*) were each clustered into three groups (Table 4).

Figure 1 shows how the variables in each group were arranged. The groups of clusters I, II, and III are shown in blue, yellow, and red color, respectively, although Fig. 1a) has only two clusters. The stars on each figure represent the centroid position for each cluster. The horizontal and vertical lines show how the range in each variable was separated. The horizontal lines separate the range of variables in the flow depth (y-axis), while the vertical lines separate

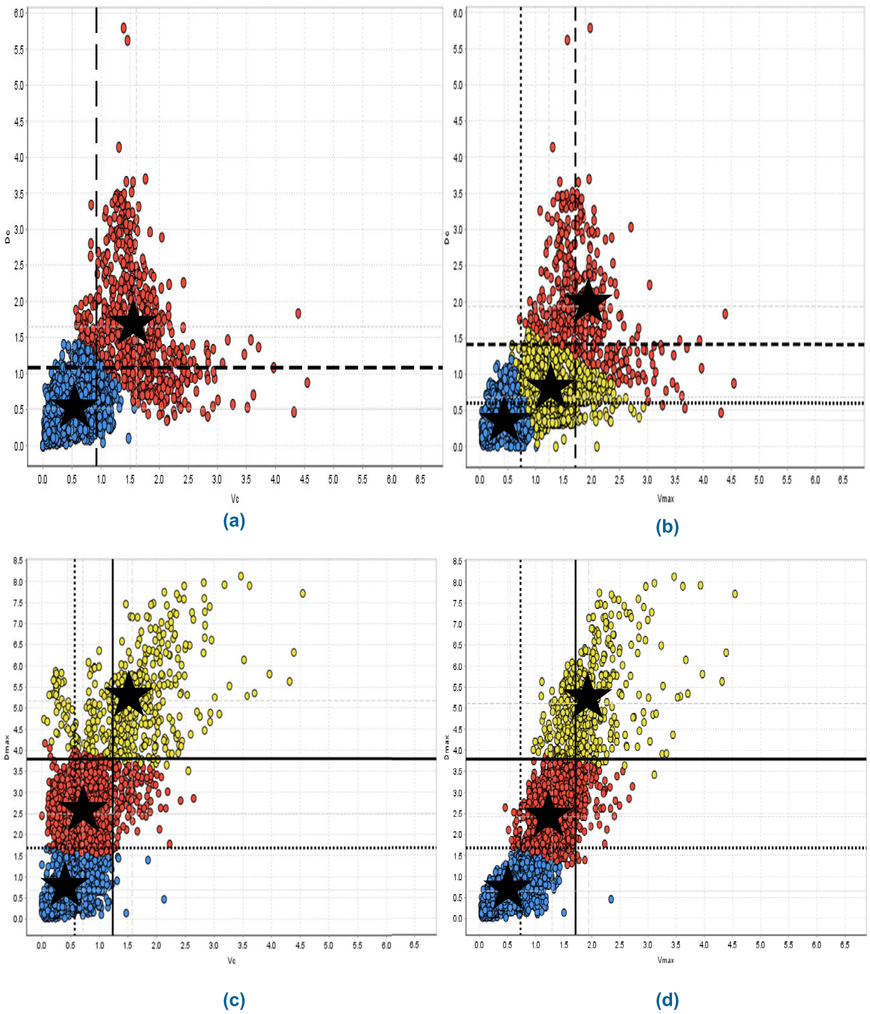


FIGURE 1. Clustering result. ***Dc&Vc*** (a); ***Dc&Vmax*** (b); ***Dmax&Vc*** (c); ***Dmax&Vmax*** (d);.

TABLE 4. Davies Bouldin index (DBI) result.

(a) *Dc* & *Vc*		(b) *Dc* & *Vmax*	
Number of clusters	DBI	Number of clusters	DBI
2	**0.711**	2	0.898
3	0.766	**3**	**0.746**
4	0.882	4	0.772
5	0.82	5	0.847
(c) *Dmax* & *Vc*		**(d) *Dmax* & *Vmax***	
Number of clusters	DBI	Number of clusters	DBI
2	0.652	2	0.613
3	**0.575**	**3**	**0.536**
4	0.683	4	0.602
5	0.76	5	0.671

the flow velocity (x-axis). The ranges of each cluster are also separated by the characteristic of the lines. The dashed line separates clusters I and II, whilethe solid line separates clusters II and III. Therefore, the dashed horizontal and vertical lines separated clusters I and II for the flow depth and flow velocity, respectively, while the solid horizontal and vertical lines separated clusters II and III for the flow depth and flow velocity, respectively.

The summary of the centroid and the range of each clustering method are also shown in Table 5. From these results, clusters I, II, and III were classified as a low, medium, and high level of variable, respectively. For example, cluster 0 from a) is classified as low level of both *Dc* and *Vc*. On the other hand, there were only two clusters for the *Dc* & *Vc* group, so cluster I and II in this case are classified as a low and high level of variable, respectively. The numerical value of the centroid of each clustering group is also shown in Table 5, along with the range of each variable in each clustering group.

IV. RESULT

In this section, the clustering analysis results are explained in Sub-section A.1 followed by the classification analysis results in Sub-section B.

TABLE 5. Centroid and range of each group of variables.

(a) *Dc* & *Vc*

Cluster number	Cluster I (Low level)		Cluster II (High level)	
Variable	*Dc* (m)	*Vc* (m/s)	*Dc* (m)	*Vc* (m/s)
Centroid	0.51	0.501	1.646	1.606
Range	$Dc < 1.08$	$Vc < 0.92$	$Dc \geq 1.08$	$Vc \geq 0.92$

(b) *Dc* & *Vmax*

Cluster number	Cluster I (Low level)		Cluster II (Medium level)		Cluster III (High level)	
Variables	*Dc* (m)	*Vmax* (m/s)	*Dc* (m)	*Vmax* (m/s)	*Dc* (m)	*Vmax* (m/s)
Centroid	0.365	0.434	0.682	1.239	1.937	1.89
Range	$Dc < 0.6$	$Vmax < 0.73$	$0.6 \leq Dc < 1.41$	$0.73 \leq Vmax < 1.71$	$Dc \geq 1.41$	$Vmax \geq 1.71$

(c) *Dmax* & *Vc*

Cluster number	Cluster I (Low level)		Cluster II (Medium level)		Cluster III (High level)	
Variables	*Dmax* (m)	*Vc* (m/s)	*Dmax* (m)	*Vc* (m/s)	*Dmax* (m)	*Vc* (m/s)
Centroid	0.682	0.444	2.484	0.711	5.169	1.577
Range	$Dmax < 1.68$	$Vc < 0.57$	$1.68 \leq Dmax < 3.79$	$0.57 \leq Vc < 1.24$	$Dmax \geq 3.79$	$Vc \geq 1.24$

(d) *Dmax* & *Vmax*

Cluster number	Cluster I (Low level)		Cluster II (Medium level)		Cluster III (High level)	
Variables	*Dmax* (m)	*Vmax* (m/s)	*Dmax* (m)	*Vmax* (m/s)	*Dmax* (m)	*Vmax* (m/s)
Centroid	0.657	0.546	2.422	1.291	5.116	1.943
Range	$Dmax < 1.68$	$Vmax < 0.73$	$1.68 \leq Dmax < 3.79$	$0.73 \leq Vmax < 1.71$	$Dmax \geq 3.79$	$Vmax \geq 1.71$

TABLE 6. Result of the clustering method.

	Dc & Vc	Dc & Vmax[b]	Dmax & Vc	Dmax & Vmax
Correlation coefficient[a]	0.568	0.552	0.685	0.860
Most impact variable	Vc	Vmax	Dmax	Dmax
Overall accuracy of the specified range	88.6%[c]	69.2%	63.2%	72.9%

[a]Correlation coefficient higher than 0.6 (the *Dmax* & *Vc* and *Dmax* & *Vmax* are rejected)
[b]The *Dc* & *Vmax* group has the highest accuracy for the decision tree
[c]The *Dc* & *Vc* group has the highest accuracy of the specified range

A. CLUSTERING ANALYSIS

In Section 3, the K-mean algorithm was used for clustering each variable group to become categorical data. However, before taking each group of variables to perform the classification analysis, correlation between the two groups of variables was tested for to avoid a multicollinearity problem. The result of the clustering and correlation analyses are shown in Table 6. To classify the damage level using two variables per group, the correlation coefficient needs to be less than 0.6 [49] to avoid a potential multicollinearity problem. Therefore, the *Dmax* & *Vc* and *Dmax* & *Vmax* groups should not be used for predicting the damage level. From the root node for each model in the DT analysis, *Vc* has more impact on the damage level than *Dc*, and *Vmax* has more impact on the damage level than *Dc*. Therefore, the flow velocity is more influential on the damage level than the flow depth. The result from this research is in accord with previous studies [2], [16] that also stated that the *Vc* was more influential on the damage level than the *Dc*.

The accuracy results on the test data for each group of variables also showed a good model performance in estimating the true damage levels, with 64.0% accuracy of the *Dc* & *Vc*, group, (i.e., 64.0% of the testing data set of the variable *Dc* & *Vc* group can predict the true damage levels). In contrast, the *Dc* & *Vmax* group had an accuracy rate of 84.7%.Finally, the accuracy of the specified range is also provided. The accuracy of the specified range is the accuracy for the clustered data that is actually in the specified range based on Table 5. The group of *Dc* & *Vc* had an accuracy range of 88.6, while that for the *Dc* & *Vc* group can cover the true cluster by 88.60%.However, the *Dc* & *Vmax* group had an accuracy for the specified range of only 69.2%.

B. CLASSIFICATION ANALYSIS

Before performing the classification analysis, the damage level based on MLIT criteria were clustered into the two groups of (1) repairable and (2) non-repairable damage levels. The condition of this criteria follows [14] by stating that the condition for damage levels above level 5 (collapse damage level) as non-repairable or a great cost for retrofitting. In contrast, damage levels below level 5 are considered as repairable or possible to be used. After that, the best setting for each classification DTRA was evaluated using the built-''Optimize Parameters'' (Grid) operator. The set-up criteria for each algorithm were well-defined and gave good results (*see* Table 7).

TABLE 7. Parameter settings for three classification DTRAs.

Parameter setting	DT	RF	GBT
Minimal leaf size	21	21	-
Minimal size for split	51	60	-
Maximal depth	20	20	20
Number of trees	1	81	24
Voting strategy	-	Majority vote	-

The minimal leaf size, minimal size for split, maximal depth, number of trees, and voting strategies were selected to be the focusing parameters. Previously, the parameter for minimal size for split and minimal leaf size were selected by a trial and error method [50]. However, in this research, the Optimize Parameters (Grid) function was used in order to find the optimal parameters setting. Then, before performing the classification analysis, the ''cross validation operator'' was applied. Using this operator, the number of k or the number of subsets of the data were set to be 10 folds due to the independent size of the training example, which can also reduce the bias problem [51], [52]. After that, each group of variables were analyzed using the DT, RF, and GBT algorithms.

The result of the DTRAs is illustrated in Fig. 2, which shows an example of the DT classification analysis based on the group of *Dmax & Vmax* after transforming to categorical data using the K-Mean clustering algorithm. Based on this, *Dmax* represented the most important variable, shown by being the top root node of the DT. In contrast, *Vmax* had a lower impact for classifying the damage level, with the damage level represented as the ending node of the decision tree. The result of the damage level was generated by the decision node of the decision tree, which was represented by the level of *Dmax* and *Vmax*. The result of the classification analyses is shown in Table 8, where each variable group was analyzed by the three classification DTRAs. The *Dmax & Vc* group had the highest classification accuracy when applying the DT or RF classification algorithm, but when applying the GBT classification algorithm the *Dmax & Vmax* group performed better than the *Dmax & Vc* group.

V. DISCUSSION

As previously reported (e.g., [1], [3], the *Dc*, *Vc*, *Dmax*, and *Vmax* were found to be the significant explanatory variables for predicting the damage levels to buildings. The *Dc* and *Vc* were used as the criterion variables compared with *Dmax* and *Vmax* to assess the building damage level from the 2011 GEJET. However, this study found that using the maximum variables might provide an underestimated level of the building damage. Thus, the critical variables were used

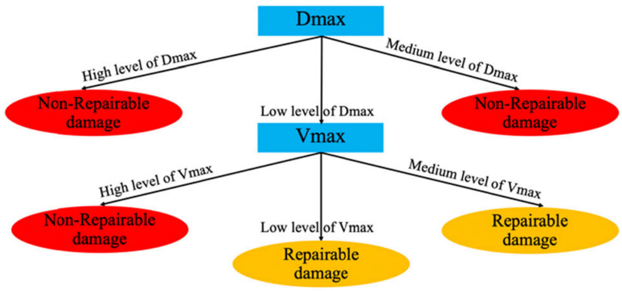


FIGURE 2. Result of the DT algorithm-based analysis.

TABLE 8. Classification accuracy of the three classification DTRAs.

Parameter group	DT	RF	GBT
Dc & Vc	64.44 ± 0.81%	64.44 ± 0.81%	59.74 ± 7.46%
Dc &V max	79.35 ± 0.98%	79.35 ± 0.98%	78.90 ± 0.91%
Dmax & Vc	89.56 ± 0.55%	89.56 ± 0.55%	86.09 ± 5.34%
Dmax & Vmax	89.55 ± 0.53%	89.55 ± 0.53%	89.55 ± 0.53%

as described in [48], where *Dc* together with *Vc* was used as a focusing variable to assess the damage level without considering any other variables. In this study, the analysis extended what previous researches have done, where the various combinations of critical and maximum variables were taken into consideration and the DTRAs were implemented as a new methodology to find the most appropriate variables to predict the damage level from a tsunami (specifically the 2011 GEJET). Table 6 shows that the group of variables most appropriate and reasonable to use in the prediction were the *Dc* & *Vmax* and *Dc* & *Vc* groups. According to the correlation result, it is obvious that the group of variables with a strong correlation must be eliminated and not taken into consideration. Even though *Dmax* can easily be obtained with high accuracy, the *Dmax* & *Vc* and *Dmax* & *Vmax* groups had a correlation coefficient of more than 0.6, a moderately strong correlation [49], and so these two groups were eliminated from this study since the variables were not independent from each other.

Comparing between the *Dc* & *Vmax* and the *Dc* & *Vc* groups, the focusing point is the accuracy rate from the DTRAs and the overall accuracy of the specifying range. Interestingly, if the *Dc* & *Vmax* group was selected, the prediction accuracy would be higher than with the *Dc* & *Vc* group.In practice, it means that selecting the *Dc* & *Vmax* group will give a more precise prediction than the *Dc* & *Vc* group.Nevertheless, if looking at the overall accuracy of the specifying range, the lower accuracy rate in the *Dc* & *Vmax* group indicates that researchers have to spend more time classifying the damage level than with the *Dc* & *Vc* group.

The classification accuracy for the three classification DTRAs also indicated that the critical variable (*Vc*) and maximum variable (*Dmax*) should be used together to classify the damage levels of buildings when using the DT and RF algorithms. However, the classification accuracy for GBT when using *Dmax* & *Vc* group was less than the classification accuracy for GBT when using the *Dmax* & *Vmax* group. Moreover, the classification accuracies when using the *Dmax* & *Vmax* group for the three classification DTRAs were almost equal to that when using the *Dmax & Vc* group. Therefore, using both maximum variables can obtain better results for classifying the damage level. The results of this research are also similar to those of [25], [53], who stated that using the maximum variables could be used for classifying the building damage levels from a tsunami.

VI. CONCLUSION

In this research, three DTRAs (DT, RF, and GBT) were used for classifying the building damage levels from a tsunami (using the 2011 GEJET) based on the maximum and critical variables of the tsunami. The group of critical and maximum variables were classified into the four groups of *Dc & Vc*, *Dc & Vmax*, *Dmax & Vc,* and *Dmax & Vmax* to evaluate whether the critical variable of the tsunami can create the same impact as the maximum variable. For the data preprocessing, the critical and maximum variable were clustered by K-mean algorithm, which separated the variable in each group is separated into the three categories of high, medium, and low levels, except for the *Dc & Vc* group that had only two categories (i.e., high and low levels). After classifying the building damage levels, the result revealed that using the maximum value of a tsunami can create a better classification accuracy based on the use of the three different DTRAs.

Nonetheless, it is crucial to understand that this study still has some limitations that can affect further predictions First, based on Fig. 1, the results of the clustering analysis do not clearly separate the level of each variable. So, the results of the DTRAs might not be accurate. The building samples in the dataset were only collected from residential buildings made from wooden materials. Therefore, if future analysis considers other types of buildings and includes other materials apart from wood, the predictive capability of the model might be different. In addition, the damage level can be influenced by other variables, such as the volume of the building and water, distance from the shoreline, and the environment around the area. Therefore, if these parameters are available, future research can include these variables in the analysis process to identify the new influence variables and enhance the accuracy of the prediction.

The results of this study can be used to support academic research, and industrial and government practices. In terms of tsunami research, this study used three DTRAs (DT, RF, and GBT) as a new approach to observe and identify the influential variables based on the updated dataset from the 2011 GEJET, and the appropriate group of variables were obtained by correlation analysis. For future studies in disaster, when these variables are available to predict the damage level, the analysis model can use these results as a comparison to check for improvements in the prediction. However, correlation between the variables is of concern. This also depends on the methodology, where this study used the DTRAs and the correlation was not a problem in this analysis. On the other hand, if the prediction is performed based on other statistical techniques, such as regression, the effects of correlation between variables should be thoroughly investigated. In addition, other machine learning techniques in addition to DTRAs, such as SVM and "*Artificial Neural Network"* (ANN), can be used to perform analyses in further studies. Additionally, the recent DTRAs, such as Forest by Continuously Excluding Root Node (Forest CERN) and Forest by Penalizing Attributes (Forest PA) can be considered in the further analysis if there are more attributes to be included. Apart from this, the researchers in disaster prevention organization, engineers in construction companies, policy makers in an insurance business, and government agencies can use such findings as a criterion for their decision-making process.

REFERENCES

[1] N. Leelawat, A. Suppasri, I. Charvet, and F. Imamura, "Building damage from the 2011 great East Japan tsunami: Quantitative assessment of influential factors," *Natural Hazards*, vol. 73, no. 2, pp. 449–471, Sep. 2014.

[2] A. Suppasri, K. Pakoksung, I. Charvet, C. T. Chua, N. Takahashi, T. Ornthammarath, P. Latcharote, N. Leelawat, and F. Imamura, "Load-resistance analysis: An alternative approach to tsunami damage assessment applied to the 2011 great East Japan tsunami," *Natural Hazards Earth Syst. Sci.*, vol. 19, no. 8, pp. 1807–1822, Aug. 2019.

[3] W. K. T. Saengtubtim, N. Wisittiwong, J. Tang, N. Leelawat, A. Suppasri, K. Pakaksung, F. Imamura, N. Takahashi, and I. Charvet, "Building damage analysis from the 2011 Great East Japan tsunami," in *Proc. Int. Conf. Eng. Natural Sci.-Bangkok (ICENS Bangkok)*, Bangkok, Thailand, 2019, pp. 33–39.

[4] A. Suppasri, I. Charvet, J. Macabuag, T. Rossetto, N. Leelawat, P. Latcharote, and F. Imamura, "Building damage assessment and implications for future tsunami fragility estimations," in *Handbook Coastal Disaster Mitigation for Engineers Planners*. Amsterdam, The Netherlands: Elsevier, 2015, pp. 147–178.

[5] P. Latcharote, A. Suppasri, A. Yamashita, B. Adriano, S. Koshimura, Y. Kai, and F. Imamura, "Possible failure mechanism of buildings overturned during the 2011 great East Japan tsunami in the town of Onagawa," *Frontiers Built Environ.*, vol. 3, p. 16, Mar. 2017.

[6] J. Macabuag, "Tsunami damage prediction for buildings: Development of methods for empirical and analytical fragility function derivation," Ph.D. dissertation, Dept. Homeland Secur., UCL, London, U.K., 2018.

[7] *Hazus Tsunami Model Technical Guidance*, FEMA, Washington, DC, USA, 2017.

[8] Ministry of Land. Infrastructure. Transport and Tourism (MLIT). *Survey of Tsunami Damage Condition*. Accessed: Aug. 25, 2020. [Online]. Available: https://www.mlit.go.jp/toshi/toshi-hukkou-arkaibu.html

[9] S. Okada and N. Takai, "Classifications of structural types and damage patterns of buildings for earthquake field investigation," *J. Struct. Construct. Eng., Trans. AIJ*, vol. 64, no. 524, pp. 65–72, 1999.

[10] G. Grünthal, "European macroseismic scale 1998," in *Proc. Eur. Seismol. Commission (ESC)*, 1998.

[11] N. Leelawat, A. Suppasri, O. Murao, and F. Imamura, "A study on the influential factors on building damage in Sri Lanka during the 2004 Indian ocean tsunami," *J. Earthq. Tsunami*, vol. 10, no. 2, Jun. 2016, Art. no. 1640001.

[12] O. Murao and H. Nakazato, "Recovery curves for housing reconstruction in Sri Lanka after the 2004 Indian Ocean Tsunami," *J. Earthq. Tsunami*, vol. 4, no. 2, pp. 51–60, Jun. 2010.

[13] O. Murao and H. Nakazato, "Vulnerability functions for buildings based on damage survey data in Sri Lanka after the 2004 Indian Ocean tsunami," in *Proc. Int. Conf. Sustain. Built Environ.*, Kandy, Sri Lanka, 2010, pp. 371–378.

[14] A. Suppasri, E. Mas, I. Charvet, R. Gunasekera, K. Imai, Y. Fukutani, Y. Abe, and F. Imamura, "Building damage characteristics based on surveyed data and fragility curves of the 2011 great East Japan tsunami," *Natural Hazards*, vol. 66, no. 2, pp. 319–341, Mar. 2013.

[15] I. Charvet, I. Ioannou, T. Rossetto, A. Suppasri, and F. Imamura, "Empirical fragility assessment of buildings affected by the 2011 great East Japan tsunami using improved statistical models," *Natural Hazards*, vol. 73, no. 2, pp. 951–973, Sep. 2014.

[16] A. Suppasri, K. Pakoksung, I. Charvet, N. Takahashi, T. Ornthammarath, and F. Imamura, "New building damage assessment method based on numerical tsunami simulation and analytical force estimation: Case study of the 2011 Tohoku tsunami," *J. Jpn. Soc. Civil Eng. B2, Coastal Eng.*, vol. 74, no. 2, pp. I-451–I-456, 2018.

[17] H. Kreibich, K. Piroth, I. Seifert, H. Maiwald, U. Kunert, J. Schwarz, B. Merz, and A. H. Thieken, "Is flow velocity a significant parameter in flood damage modelling?" *Natural Hazards Earth Syst. Sci.*, vol. 9, no. 5, pp. 1679–1692, Oct. 2009.

[18] S. Koshimura, T. Oie, H. Yanagisawa, and F. Imamura, "Developing fragility functions for tsunami damage estimation using numerical model and post-tsunami data from Banda Aceh, Indonesia," *Coastal Eng. J.*, vol. 51, no. 3, pp. 243–273, Sep. 2009.

[19] L. R. Iverson and A. M. Prasad, "Using landscape analysis to assess and model tsunami damage in Aceh province, Sumatra," *Landscape Ecol.*, vol. 22, no. 3, pp. 323–331, Feb. 2007.

[20] K. Harada and F. Imamura, "Effects of coastal forest on tsunami hazard mitigation—A preliminary investigation," in *Tsunamis*. Dordrecht, The Netherlands: Springer, 2005, pp. 279–292.

[21] C. Petrone, T. Rossetto, and K. Goda, "Fragility assessment of a RC structure under tsunami actions via nonlinear static and dynamic analyses," *Eng. Struct.*, vol. 136, pp. 36–53, Apr. 2017.

[22] S. Koshimura, M. Matsuoka, and S. Kayaba, "Tsunami hazard and structural damage inferred from the numerical model, aerial photos and SAR imageries," in *Proc. 7th Int. Workshop Remote Sens. Post Disaster Response*, 2009, pp. 22–23.

[23] R. De Risi, K. Goda, T. Yasuda, and N. Mori, "Is flow velocity important in tsunami empirical fragility modeling?" *Earth-Sci. Rev.*, vol. 166, pp. 64–82, Mar. 2017.

[24] N. Leelawat, A. Suppasri, I. Charvet, T. Kimura, D. Sugawara, and F. Imamura, "A study on influential factors on building damage in Kesennuma, Japan from the 2011 great east Japan tsunami," *Eng. J.*, vol. 19, no. 3, pp. 105–116, Jun. 2015.

[25] I. Charvet, A. Suppasri, H. Kimura, D. Sugawara, and F. Imamura, "A multivariate generalized linear tsunami fragility model for Kesennuma city based on maximum flow depths, velocities and debris impact, with evaluation of predictive accuracy," *Natural Hazards*, vol. 79, no. 3, pp. 2073–2099, Dec. 2015.

[26] J. Stolle, N. Goseberg, I. Nistor, and E. Petriu, "Debris impact forces on flexible structures in extreme hydrodynamic conditions," *J. Fluids Struct.*, vol. 84, pp. 391–407, Jan. 2019.

[27] M. Bikçe, M. Örnek, and Ö. F. Cansız, "The effect of buoyancy force on structural damage: A case study," *Eng. Failure Anal.*, vol. 92, pp. 553–565, Oct. 2018.

[28] S. Koshimura and S. Kayaba, "Tsunami fragility inferred from the 1993 Hokkaido Nansei-oki earthquake tsunami disaster," *J. Jpn. Assoc. Earthq. Eng.*, vol. 10, no. 3, pp. 87–101, 2010.

[29] E. Mas, S. Koshimura, A. Suppasri, M. Matsuoka, M. Matsuyama, T. Yoshii, C. Jimenez, F. Yamazaki, and F. Imamura, "Developing tsunami fragility curves using remote sensing and survey data of the 2010 Chilean tsunami in Dichato," *Natural Hazards Earth Syst. Sci.*, vol. 12, no. 8, pp. 2689–2697, Aug. 2012.

[30] I. Charvet, A. Suppasri, and F. Imamura, "Empirical fragility analysis of building damage caused by the 2011 Great East Japan tsunami in Ishinomaki city using ordinal regression, and influence of key geographical features," *Stochastic Environ. Res. Risk Assessment*, vol. 28, no. 7, pp. 1853–1867, Oct. 2014.

[31] A. Suppasri, S. Koshimura, K. Imai, E. Mas, H. Gokon, A. Muhari, and F. Imamura, "Damage characteristic and field survey of the 2011 Great East Japan tsunami in Miyagi prefecture," *Coast. Eng. J.*, vol. 54, no. 1, 2012, Art. no. 1250005.

[32] M. Hofmann and R. Klinkenberg, *RapidMiner: Data Mining Use Cases and Business Analytics Applications*. Boca Raton, FL, USA: CRC Press, 2016.

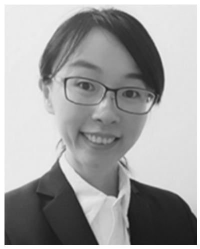
[33] M. A. Friedl and C. E. Brodley, "Decision tree classification of land cover from remotely sensed data," *Remote Sens. Environ.*, vol. 61, no. 3, pp. 399–409, Sep. 1997.

[34] M. C. Tu, D. Shin, and D. Shin, "Effective diagnosis of heart disease through bagging approach," in *Proc. 2nd Int. Conf. Biomed. Eng. Informat.*, 2009, pp. 1–4.

[35] M. N. Anyanwu and S. G. Shiva, "Comparative analysis of serial decision tree classification algorithms," *Int. J. Comput. Sci. Secur.*, vol. 3, no. 3, pp. 230–240, 2009.
[36] L. Breiman, "Random forests," Univ. Californai, Berkeley, Berkeley, CA, USA, Tech. Rep. TR567, 1999.
[37] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
[38] M. S. Alam and S. T. Vuong, "Random forest classification for detecting Android malware," in *Proc. IEEE Int. Conf. Green Comput. Commun., IEEE Internet Things IEEE Cyber, Phys. Social Comput.*, Aug. 2013, pp. 663–669.
[39] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[40] R. Díaz-Uriarte and S. A. De Andres, "Gene selection and classification of microarray data using random forest," *BMC Bioinform.*, vol. 7, no. 1, p. 3, 2006.
[41] A. V. Lebedev, E. Westman, G. J. P. Van Westen, M. G. Kramberger, A. Lundervold, D. Aarsland, H. Soininen, I. Kłoszewska, P. Mecocci, M. Tsolaki, and B. Vellas, "Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness," *NeuroImage, Clin.*, vol. 6, pp. 115–125, Jan. 2014.
[42] R. E. Schapire, Y. Freund, P. Bartlett, and W. S. Lee, "Boosting the margin: A new explanation for the effectiveness of voting methods," *Ann. Statist.*, vol. 26, no. 5, pp. 1651–1686, Oct. 1998.
[43] G. Rushin, C. Stancil, M. Sun, S. Adams, and P. Beling, "Horse race analysis in credit card fraud—Deep learning, logistic regression, and gradient boosted tree," in *Proc. Syst. Inf. Eng. Design Symp. (SIEDS)*, Apr. 2017, pp. 117–121.
[44] N. Mori, T. Takahashi, and 2011 Tohoku Earthquake Tsunami Joint Survey Group, "Nationwide post event survey and analysis of the 2011 Tohoku earthquake tsunami," *Coast. Eng. J.*, vol. 54, no. 1, 2012, Art. no. 1250001.
[45] H. Lim and S. W. Hoag, "Plasticizer effects on physical–mechanical properties of solvent cast Soluplus films," *AAPS PharmSciTech*, vol. 14, no. 3, pp. 903–910, Sep. 2013.
[46] N. Singh and D. Singh, "Performance evaluation of k-means and heirarichal clustering in terms of accuracy and running time," *Int. J. Comput. Sci. Inform. Tech.*, vol. 3, no. 3, pp. 4119–4121, 2012.
[47] J. Jamesmanoharan, S. H. Ganesh, M. L. P. Felciah, and A. K. Shafreenbanu, "Discovering students' academic performance based on GPA using K-means clustering algorithm," in *Proc. World Congr. Comput. Commun. Technol.*, Feb. 2014, pp. 200–202.
[48] J.-C. Lamirel, N. Dugue, and P. Cuxac, "New efficient clustering quality indexes," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 3649–3657.
[49] Y. H. Chan, "Biostatistics 104: Correlational analysis," *Singap. Med. J.*, vol. 44, no. 12, pp. 614–619, 2003.
[50] S. R. Dindarloo and E. Siami-Irdemoosa, "Maximum surface settlement based classification of shallow tunnels in soft ground," *Tunnelling Underground Space Technol.*, vol. 49, pp. 320–327, Jun. 2015.
[51] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Joint Conf. Artif. Intell.*, Montreal, QC, Canada, vol. 14, 1995, pp. 1137–1145.
[52] J.-H. Kim, "Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap," *Comput. Statist. Data Anal.*, vol. 53, no. 11, pp. 3735–3745, Sep. 2009.
[53] H. Matsutomi, K. Okamoto, and K. Harada, "Inundation flow velocity of tsunami on land and its practical use," *Coastal Eng. Proc.*, vol. 1, no. 32, p. 5, Jan. 2011.

NATT LEELAWAT (Senior Member, IEEE) received the B.Sc. degree (Hons.) in information technology from the Sirindhorn International Institute of Technology, Thammasat University, Thailand, in 2007, and the M.Eng. and D.Eng. degrees in industrial engineering and management from the Tokyo Institute of Technology, Japan, in 2013 and 2016, respectively.

He was a System Analyst with the Bank of Thailand; and an Assistant Professor with Tohoku University, Japan. He is currently an Assistant Professor with the Department of Industrial Engineering, Faculty of Engineering, Chulalongkorn University, Thailand. He is also a Director of the Risk and Management Program, Graduate School, an Assistant Dean of the Faculty of Engineering, and the Head of Disaster and Risk Management Information Systems Research Group, Chulalongkorn University. His research interests include management information systems, disaster and risk management, and business continuity management. His awards and honors include the ICA-SYMP 2021 Best Paper Award, the Tokyo Tech Alumni Association (Thailand Chapter)'s Young Outstanding Alumni Award 2019, the SakIntania's Young Lecturer Award 2018, and the SUSTAIN 2015 Best Paper Award.



JING TANG received the B.Mgmt. degree in industrial engineering and the B.Eng. degree in computer science and technology from Xi'an Jiaotong University, China, in 2008, and the M.Eng. and D.Eng. degrees in industrial engineering and management from the Tokyo Institute of Technology, Japan, in 2010 and 2013, respectively. She was a Lecturer with the Sirindhorn International Institute of Technology, Thammasat University, Thailand. She is currently a Lecturer of Robotic and Artificial Engineering Program and Information and Communication Engineering Program with the International School of Engineering, Faculty of Engineering, Chulalongkorn University, Thailand. Her research interests include data science and data analytics, business intelligence and artificial intelligence, business process management, business process outsourcing, and simulation and modeling.

WANIT TREERANURAT received the B.Sc. degree (Hons.) in engineering management from the Sirindhorn International Institute of Technology, Thammasat University, Thailand, in 2019, where he is currently pursuing the degree major in logistic and supply chain systems engineering. He is currently a Solution Consultant with Stream I.T. Company Ltd., Thailand.

KUMPOL SAENGTABTIM received the B.Sc. degree (Hons.) in engineering management from the Sirindhorn International Institute of Technology, Thammasat University, Thailand, in 2019. He is currently pursuing the degree with the Department of Industrial Engineering, Faculty of Engineering, Chulalongkorn University, Thailand.

NARUNPORN WISITTIWONG received the B.Sc. degree in engineering management from the Sirindhorn International Institute of Technology, Thammasat University, Thailand, in 2019. From 2019 to 2020, she was a Consulting Development Analyst with Accenture Solutions Company Ltd., Thailand.

ANAWAT SUPPASRI received the B.Eng. degree in civil engineering from Chulalongkorn University, in 2005, the M.Eng. degree in water engineering from the Asian Institute of Technology, Thailand, in 2007, and the Ph.D. degree in tsunami engineering from Tohoku University, Japan, in 2010.

From 2010 to 2012, he was a Postdoctoral Research Fellow with the Disaster Control Research Center, Tohoku University. He is currently an Associate Professor with the International Research Institute of Disaster Science, Tohoku University. His research interests include tsunami numerical modeling, tsunami generation mechanism, tsunami damage field survey, tsunami vulnerability, tsunami hazard and risk evaluation, and disaster prevention education. His awards and honors include the Coastal Engineering Journal Citation Award, the Nakasone Yasuhiro Award, Hazard 2000 Award, the JAMSTEC Nakanishi Award, and the Hodaka Prize. He serves as an Editor for many journals, including *Coastal Engineering Journal*, *International Journal of Disaster Risk Reduction*, *Geoscience Letters*, and *Geosciences*.

KWANCHAI PAKOKSUNG received the B.Eng. degree in civil engineering from the Rajamangala Institute of Technology, the M.Eng. degree in water resources engineering from Chulalongkorn University, Thailand, and the D.Eng. degree in infrastructure engineering from the Kochi University of Technology, Japan. He is currently a Researcher with the International Research Institute of Disaster Science, Tohoku University, Japan. His research interests include tsunami numerical modeling, multi-layer flow simulation, landslide modeling, economic losses of disaster, flood numerical modeling, remote sensing data for flood simulation, flood hazard and risk evaluation, and distributed hydrological modeling.

FUMIHIKO IMAMURA received the B.S. degree in civil engineering, and the M.Eng. and D.Eng. degrees in coastal engineering from Tohoku University, Japan. He is a currently a Professor in tsunami engineering with Tohoku University and a Director of the International Institute of Disaster Science, Tohoku University. His research interests include tsunami engineering, numerical simulation for tsunamis, real time tsunami warning, disaster mitigation technology, and human response for information and evacuation. He serves as a Council Member for the International Society for the Prevention and Mitigation of Natural Hazards, an Associate Member for the Science Council of Japan, a Member of CIP Committee on International Participation, and a President of the Japan Society for Natural Disaster Science.

NORIYUKI TAKAHASHI received the B.Eng., M.Eng., and Ph.D. degrees in architectural engineering from The University of Tokyo, Japan, in 2000, 2002, and 2005, respectively. He was an Assistant Professor with the Institute of Industrial Science, The University of Tokyo, from 2005 to 2013. Since 2013, he has been an Associate Professor with Tohoku University. His awards and honors include the Japan Concrete Institute Award (Encouragement Award), in 2004, and the Technology Award in Tohoku branch, in 2020. He serves as an Editor for journals, including *Journal of Structural Engineering*.

INGRID CHARVET received the B.S. degree in hydrographic surveying from the Conservatoire National des Arts et Metiers/CNAM, France, in 2006, and the M.Sc. degree in geophysical hazards, the Ph.D. degree in tsunami engineering, and the M.Sc. degree (Hons.) in statistical data science from UCL, U.K., in 2007, 2011, and 2019, respectively. She was a Statistical Data Scientist with UCL, a Lead Risk Modeler with RMS, U.K., an Honorary Researcher with UCL, and a Research Fellow with Tohoku University, Japan. She is currently a Quantitative Developer with Praedicat, Inc., U.K. Her research interests include latent liability risk modeling, building vulnerability analysis, stochastic damage modeling, and sports injury risk analysis.