A framework of Multi Linear Regression based on Fuzzy Theory and Situation Awareness and its application to Beach Risk Assessment

Gun-Yoon Shin[1], Sung-Sam Hong[1], Dong-Wook Kim[1], Cheol-Hun Hwang[1], Myung-Mook Han*[1], Hwayoung Kim[2] and Young jae Kim[3]

[1] Graduate School of IT Convergence Engineering Dept., Gachon University, Seongnam, Republic of Korea
[e-mail: tlsrjsdbs@gmail.com, sungsamhong@gmail.com, kog7306@naver.com, qewqsa@naver.com, mmhan@gachon.ac.kr]

[2] Division of Maritime Transportation, Mokpo national Maritime University, Mokpo, Republic of Korea
[e-mail: hwayoung@mmu.ac.kr]

[3] Ji In system, Seoul, Republic of Korea
[e-mail: kimyi@ji-in.com]

*Corresponding author: Myung-Mook Han

Abstract

Beaches have many risk factors that cause various accidents, such as drifting and drowning, these accidents have many risk factors. To analyze them, in this paper, we identify beach risk factors, and define the criteria and correlation for each risk factor. Then, we generate new risk factors based on Fuzzy theory, and define Situation Awareness for each time. Finally, we propose a beach risk assessment and prediction model based on linear regression using the calculated risk result and pre-defined risk factors. We use national public data of the Korea Meteorological Administration (KMA), and the Korea Hydrographic and Oceanographic Agency (KHOA). The results of the experiment showed the prediction accuracy of beach risk to be 0.90%, and the prediction accuracy of drifting and drowning accidents to be 0.89% and 0.86%, respectively. Also, through factor correlation analysis and risk factor assessment, the influence of each of the factors on beach risk can be confirmed. In conclusion, we confirmed that our proposed model can assess and predict beach risks.

1. Introduction

Many tourists visit beaches, and their numbers have been increasing every year. As the number of tourists to beaches increases, accidents associated with beaches are also increasing. However, the accidents occurring at the beach have not been clearly defined, and the associated risk factors beach accident are not obvious [1].

It is difficult to regard accidents occurring at a beach as being caused by one simple factor. Rather, they are regarded as a result of compositing interactions of environmental factors at the beach that includes water temperature, air temperature, wind direction, wind speed, current speed, current direction, and so on. Personal factors of individual that includes personal carelessness, drinking, age, disease, etc. [2]. For example, in an accident associated with drifting, a tourist swimming with a tube may be floated away in an unexpected direction influenced by tide or current speed or direction or wind. Even though tourists recognize the environmental risk factor, accidents happen.

Therefore, in this paper, we study risk analysis, beach assessment, and risk prediction that could happen environmental factors occurred at a beach [3, 4]. First, perform investigation of the risk factors. Second, define risk criteria of each factor by utilizing Fuzzy theory. Third, compute beach risks by using Situation Awareness and Fuzzy theory. After carrying out all above orders, a beach risk assessment model will be defined based on Multi Linear Regression. It allow us to calculate the beach risk for each time, and find the main risk factors at the beach. Finally, We will be able to predict the risk of the beach.

The study can be summarized as follows. First, Section 2 introduces studies related to beach risk assessment. Section 3 then discusses the beach risk assessment Equation, framework, and processes. Section 4 next presents the risk assessment results for drowning and drifting, the beach risk assessment for a beach, and the results of analysis on the main risk factors according to the corresponding situation. Section 5 finally concludes the paper.

2. Related Work

Beach risk assessment can be regarded as occurring with composite interaction of various factors, and to establish this clearly, we need to check beach accident information. However, very little information about beach accidents is being collected, and it can be confirmed that even the aggregation of accidents is not performed exactly. Therefore, in order to systematically analyze and manage beach accidents, risk factors of the beach should be exactly acknowledged, and a system for the analysis, assessment, and prediction of beach risk is needed [5, 6].

Beach risk assessment means computing risks on a beach with various risk factors, and is mainly research in England and Australia, which have many beaches. Fisher et al [7]. researched the seashore environment in England and Ireland, and performed beach risk assessment investigation through a 3-step process based on scientific principle, use cases, and benchmarks. A beach risk assessment is performed by referring to the relevant institution's medical guidelines, and the risks to beaches are divided into environment and human category:

$$\text{Risk} = \text{E(Tide + Average wave height or Flow)} + \text{P(In water population + CA)} \qquad (1)$$
$$+/- \text{ UKBSAMP}$$

Risks are measured by Eq. (1) in the study, where E is energy, P is population, and UKBSAMP weight is the numerical value that is evaluated of the beach type. E is represented by 7 steps using the criteria of tides and average height of wave or tide flow, P is also represented with 7 steps using criteria of the population in water, and CA means Conflicting activity. UKBSAMP weight is calculated with beach facilities, beach activities, beach visitor information, building environment around the beach, natural environment, water quality, beach information, and so on, and divided into 14 beach types. Beach facilities determines whether there are facilities, such as beach parking lots, shower facilities, toilets and bath houses, and beach activities include beach landscape, roads, available family activity, local and nature tour and water activity; while visitor information includes number of families, children, weak or senior persons and visitors from foreign countries. Building environment includes whether there are resorts, main street or municipal region, or sea resorts, and, in natural environment, the presence of sand and mud, water quality, and beach award history.

In *NSW Government* [8], Australia developed a risk assessment model to build a safety system that can organically respond to life accidents and disasters occurring at a beach, and for risk assessment of a beach, using natural environmental features, safety facility, status of population around a beach, and status of visitors at a beach per time:

$$\text{Risk} = (\text{ABSAMP} \cdot 2) + (\text{LPR} \cdot 2) + (\text{HAIR} \cdot 2) + (\text{AR} \cdot 2) \qquad (2)$$

Equation (2) is the method of risk assessment proposed in the corresponding paper, where the Australia Beach Safety & Aquatic Management Program (ABSAMP) divides the risk level of the natural state at a beach based on criteria, Local Population Rate (LPR) is the risk by the population within 2 km of radius from the beach, HAIR is the number of visitors that engage in activity at the beach, and Access Rate (AR), means the accessibility to a beach. The calculated beach risk results are divided into 5 levels, depending on the score.

Byungmo et al. [9] provides a research assessment of about 300 beaches in Korea, and divides beaches into 9 types in consideration of wave, tide, shape of sandy beach, surrounding environment, and so on. First, they divide beach into 2 types using wave tide, and detailed classification has been performed by adding information, such as the shape of sandy beach and surrounding environment. For 9 types of classified beaches, they define the features and risk values for each beach, with the risk value range being 0 to 1, where the closer the value is to 1, the higher the risk. Their beach risk assessment framework consists of 3 steps; in step 1, the information collected through general methods of beach classification, like average height of wave, local environment information, and shape of sandy beaches, is used to perform risk identification. Accidents that can occur according to potential risk factors, main causes of risks, wave, and shapes of sandy beaches are confirmed through the risk assessment, and then risk analysis is performed:

$$\text{Risk} = S \cdot N \cdot [(1 - E) \cdot (1 - W) \cdot (1 - P) \cdot H] \cdot F \qquad (3)$$

In step 2, main risks are certainly recognized, and are clearly defined within the range that can be controlled; and, in step 3, risk assessment is performed based on the information obtained in steps 1 and 2. This third step checks whether the beach risks can be addressed at the current beach situation, what measures should be taken to resolve the risks, and what measures for the beach are established by using various methods, such as monitoring, discussion, and review. Equation (3) is a method of measuring beach risks, where S means the Korean beach risk class that is divided into 9 types, N is the number of visitors, E is the ratio of safety agents per area of the beach, W is the number of guiding signs per area of the beach, P is a safety preliminary factor of whether visitors are equipped with safety factors, H is the average height of wave, and F is the floating population per year. Each of the values is given numerical value to perform the assessment.
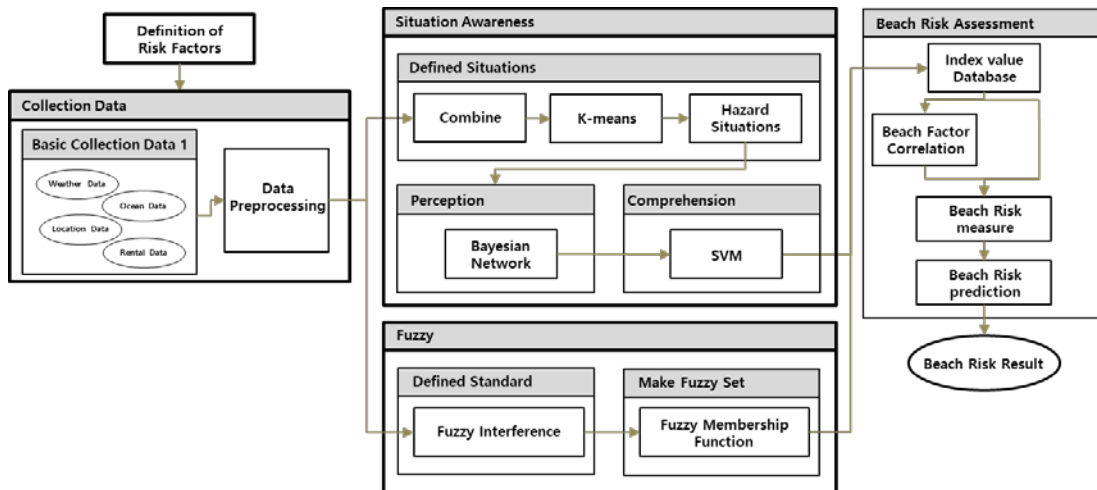
3. A framework of Beach Risk Assessment

Fig. 1. Beach risk assessment framework

In this paper, we propose a Fuzzy theory and Situation Awareness method, in order to overcome the problems in the previous method [6] of using Multi Linear Regression, in that it is difficult to identify exact risk situations by using only natural environmental factors, and establish exact risk criteria for each factor.

For the problems where it is difficult to identify exact risk situations by using natural environmental factors only, we analyse the current situation using the past situations analyzed by Situation Awareness, and then address the problem through a method of analyzing and predicting future situations. Risk situations are identified through Situation Awareness, and for situations that have not been recognized from a mental model, identification was performed through situation overlay by applying a clustering method to natural environmental factor data.

For the risk criteria of each factor, rules are generated for each factor by utilizing Fuzzy Interference, and Fuzzy Membership Function is defined through the meetings with professionals in charge of beach risk management and personnel.

The beach risk assessment model is divided into 3 steps; first, defining the risk factor criteria by using Fuzzy theory, second, identifying risk situation by using Situation Awareness,

and last, performing risk assessment and prediction by using Multi Linear Regression. **Fig. 1** illustrates the framework of a beach risk assessment model, in which all factors in public data are collected, and then only the factors are extracted that are required for beach risk assessment, and preprocessing is performed, such as data cleaning, and normalization. After that, risk definition on the factors for each time is performed through risk factor criteria defined using Fuzzy theory, and risk situation identification is performed through Situation Awareness. With the data obtained by the two methods, risk assessment and prediction are performed [10].

3.1 Fuzzy theory-based risk factor criteria definition

In this paper, Fuzzy theory is applied to establish the criteria for the factors corresponding to the natural environment among various risk factors occurring at a beach. Among 14 risk factors defined in the previous study [6], water height, depth of water, tidal current, wind direction, and wind speed factor are used. The reason is that, among the 14 risk factors, factors such as age, sex, drinking, lack of skill in swimming, and disease history are difficult to quantify, and analysis of the 14 risk factors through meetings with professionals resulted in the findings that water height, depth of water, tidal current, wind direction, and wind speed are more important than the other factors.

In order to establish the criteria for beach risk factors by applying Fuzzy theory, the concept of each of the risk factors was established through meetings with professionals, and new factors were defined combining two factors based on the criteria [11]. The reason new factors were defined by combining two types of factors is that while some accidents happen from only one factor, many accidents occur from two or more factors. So, we make new factors that combine water height and depth of water, direction and velocity of tidal flow, wind direction and wind speed. For the tidal current direction and wind direction, the beach is defined through criteria, wind direction means the direction wind blows from, while tidal current is the direction that water flows towards.

Table 1. Fuzzy Membership Function according to two types of beach accidents

Drowning Accident				Drifting Accident			
Wave height	**Depth**			**Wave height**	**Depth**		
	Small	Medium	Big		Small	Medium	Big
Small	VS	S	B	Small	VS	M	B
Medium	M	M	B	Medium	M	B	VB
Big	M	B	VB	Big	M	B	VB
Current direction	**Current velocity**			**Current direction**	**Current velocity**		
	Small	Medium	Big		Small	Medium	Big
Small	S	S	VS	Small	S	S	VS
Medium	S	M	B	Medium	M	B	VB
Big	M	M	B	Big	M	B	VB
Wind direction	**Wind velocity**			**Wind direction**	**Wind velocity**		
	Small	Medium	Big		Small	Medium	Big
Small	VS	VS	VS	Small	VS	VS	VS
Medium	S	M	M	Medium	S	M	B
Big	S	M	B	Big	M	B	VB
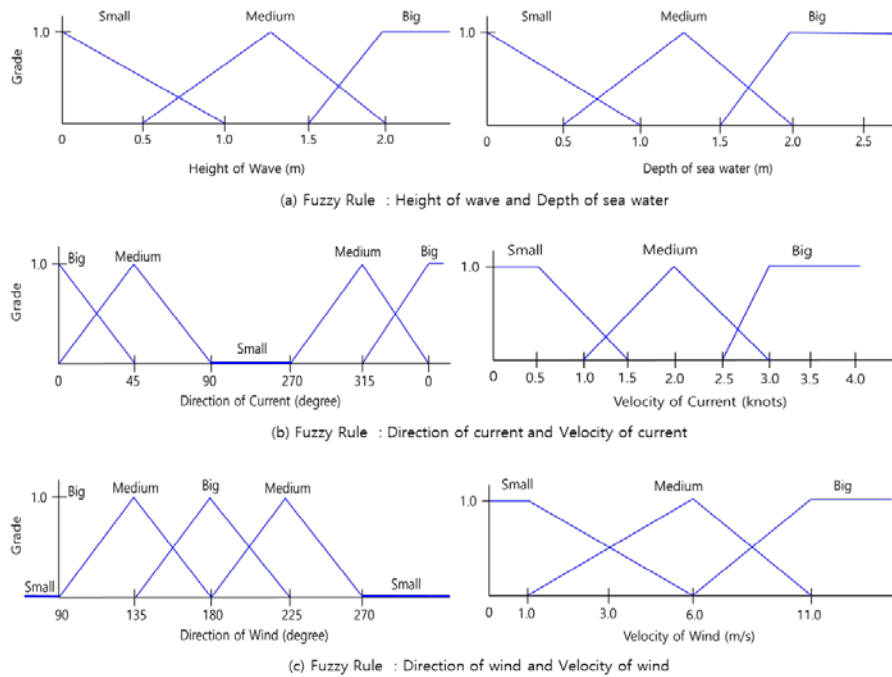
Fig. 2. Definitions of 6 types of Fuzzy Interference based on factor criteria, (a) water height and depth of water, (b) direction and velocity of tidal current, (c) risk criteria according to wind direction and wind speed. Risk goes to large means higher risk.
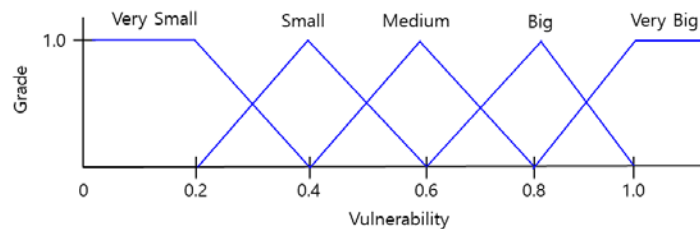
Fig. 3. Risk criteria settings for generating Fuzzy Membership Function

Criteria have been established for three new factors in drowning and drifting accidents; rules have been generated for water height, depth of water, tidal current, wind direction, and wind speed factors by utilizing Fuzzy Interference; and Fuzzy Membership Function is constructed based on the experiences of beach safety managing professionals. **Table 1** summarizes the criteria established through Fuzzy theory. Definitions are made for water height and depth of water, direction of tidal current and velocity, and wind direction and wind speed; and criteria for each factor are divided into small, medium, and big to express what influences are made. VS refers to very small risk, S is small risk, M is medium risk, B is big risk, and VB is very big risk, and this means that risk increases on moving from VS to VB.

Fig. 2 shows the risk criteria associated with the beach risk factors represented by using Fuzzy theory, and it can be confirmed that different criteria are generated for each of 6 types of factors. Also, **Fig. 3** defines new factors based on them and shows risk criteria, and Fuzzy Membership Function was defined from the risk criteria.

3.2 Situation Identification and Prediction by using Situation Awareness

Situation Awareness means predicting future situations by recognizing the current situation through environmental factors, and by understanding the current situation through data of the past [12], each environmental factor is defined as an entity, and the situation occurring through relations before and after each of the entities is defined as a status. If information on various statuses are required in parallel when interpreting the current situation based on the information, such as past situations, it is classified as dynamic situation perception. In other words, prediction of the next situation based on the information of the final situation shown currently is called Situation Awareness.
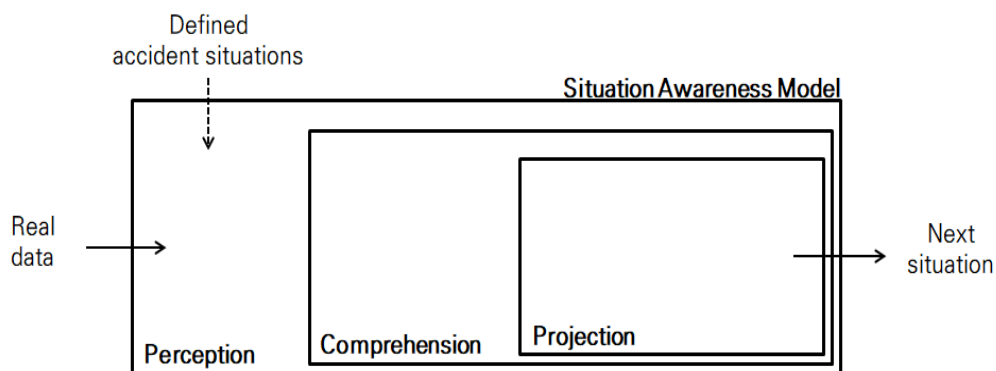

Fig. 4. Situation Awareness basic model

Fig. 4 shows a simple situation perception model, in which the step of perception is made on the current situation based on the risk situation defined through methods such as mental model or clustering, and actual data; and results are derived on the current situation, and the results are used as input data for the step of Comprehension. The step of Comprehension identifies whether the current situation is a risk situation or not, based on the results derived in the step of perception; and in the step of Projection, a future situation is decided on using the current risk situation.

In this paper, definitions have been made for accident risk situations present in the public data collected based on Situation Awareness. When multiple situations are closely related, such as in natural environments, it is difficult to identify risk situation. Therefore, in order to address these problems, clustering methods are used to identify the situations, which could not be identified from a mental model, with natural factors [13]. We use clustering to perform based on the data according to time, due to the limitations in data that can be acquired geographically. *Haidar et al.* [14] defined 10 environmental factors, and a mental model based on the factors. Then, prediction was made with the fact that relations and influences among environmental factors vary continuously over time. With reference to this, we use K-means to combine a list of defined factor features.

In situation overlay, there is the presupposition that influences from various situations and features can occur at the same time band; and these features mean that the influence of risk situation occurring due to the relations and effects among features in the situation occurring at the same time can be different. For example, in the case where the situation of growing current speed due to rising tide of beach and the situation of fast wind occur at the same time, influences of current and wind can have a different result from each other, comparing the case of opposite direction, and the case of the same direction.
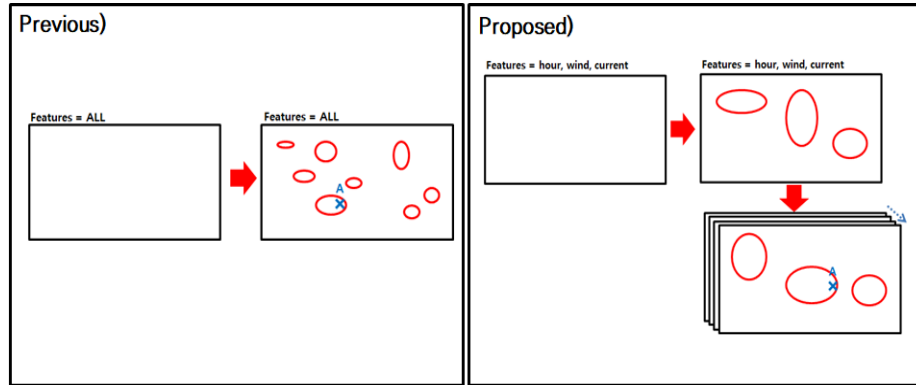
Fig. 5. Situation Overlay

In order to apply these characteristic to beach risk situation identification, a method of overlaying images is used [15], where the images mean each of the situations, and the risk probabilities of each situation derived through overlay are regarded as the extracted features. **Fig. 5** shows the above description, in which when a beach situation called 'A' occurs, the method finds in what cluster 'A' corresponds to the clustered clusters based on the list of various combined features by combining the features over time, derives the risk probability of the cluster, and then averages the derived risk probabilities to derive the risk probability of beach situation 'A'.

For the situation identification through situation overlay, defined risk situations were imported from perception step and current situation is identified based on the input actual data, then risk prediction probability is computed based on Bayesian theory for each risk situation defined.

For the situation identification through situation overlay, defined risk situations are imported from the perception step, and the present situation is identified based on the input actual data, then risk prediction probability is computed based on Bayesian theory for each risk situation defined:

$$SC = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} ||x_n - \mu_k||^2 \tag{4}$$

$$p(SC) = p(pred_{sc}|SC) \tag{5}$$

Equation (4) computes SC, while Eq. (5) computes the risk situation identification and prediction probability from each risk situation SC based on input data, where x is the actual input features; n is the number of risk features; k is the clustered clusters; and SC is the risk Situations Cluster derived through clustering by combining each of the features; $pred_{sc}$ represents the risk situation prediction results of SC; n represents the size of the risk situation cluster SC in the list of feature combinations derived from x; and m is the size of risk situation S derived from x, occurring at the same time:

$$p(S) = \frac{1}{n} \sum_{i=1}^{n} p(SC_i) \tag{6}$$

$$p(Accident) = \frac{1}{m} \sum_{i=1}^{m} p(S_i) \tag{7}$$

Equation (6) computes the averaged risk probability of all risk situation S in the same list of combined features based on the probability computed from Eq. (5). Equation (7) computes the averaged value of risk prediction probabilities of all defined risk situation S that occur at the same time. Finally, Support Vector Machine (SVM) is used to compare the average value of risk prediction probabilities and the actual accident, to determine how consistent it is.

3.3 Multi Linear Regression-based Beach Risk Assessment and Prediction

Beach risk assessment and prediction are performed with the data obtained from the two methods defined above. The status of a pre-researched actual accident is set to dependent variables, and the features obtained from Fuzzy theory and the natural factor features obtained from Situation Awareness are set to independent variables. In order to assess the risks of a beach, the risks of drowning and drifting accidents were assessed first. Since the main risk factors are different for drowning and drifting, two methods defined previously were used to generate each data for drowning and drifting accidents. In the case of Multi Linear Regression, if many features are used as independent variables, good result are not obtained. So, they need to be adjusted to an appropriate number of features. Therefore, we addressed this issue by the method of adjusting the number of features through feature correlation analysis. Feature correlation analysis was possible, since there is information, like wind direction and current direction, wind speed and current speed, and so on, that can occur similarly over time; and features included in the generated data set were constructed with continuous numerical values. According to this process, the correlation information among the features was confirmed, and unnecessary features removed. When performing correlation analysis, Pearson correlation coefficient [16] was used, and this method represents the linear correlation of two features as a parametric method, which is represented as in Eq. (8):

$$r_{x,y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \tag{8}$$

Table 2. Definition of Pearson correlation coefficient size

Range(%)	Definition
-1.0 =< r <= -0.7	Very strong negative (-) correlation
-0.7 < r <= -0.3	Strong negative (-) correlation
-0.3 < r <= -0.1	Weak negative (-) correlation
-0.1 < r <= 0.1	No correlation
0.1 < r <= 0.3	Weak positive (+) correlation
0.3 < r <= 0.7	Strong positive (+) correlation
0.7 < r <= 1.0	Very strong positive (+) correlation

where, μ_X means the average of population X, μ_Y means the average of population Y, and σ_X and σ_Y mean the standard deviations of populations X and Y, respectively. Although the size of the Pearson correlation coefficient is not absolute, generally, the ranges were set as in **Table 2** to represent the size of correlation between two features.

$$p\left(Accident_{risk\ factor}\right) = \frac{1}{m}\sum_{i=1}^{m} p(S_i)$$

$$Risk_{(Drowning\ or\ Drifting)} = \sum p\left(Accident_{risk\ factor}\right)$$

(9)

After removing unnecessary features through correlation analysis, risk assessment was performed. The obtained values were used in risk assessment from the pre-defined Fuzzy theory and Situation Awareness. All values for each factor at a certain time were computed, and risk assessment and prediction on drowning and drifting accident was performed with the values. The values for each factor were applied to Eq. (9), and a line representing the dependent variable best was found and defined; then, an assessment and prediction were made. At this step, each risk for drowning and drifting accidents was computed, and an assessment and prediction was performed. Here, $p\left(Accident_{risk\ factor}\right)$ means extracting the probability value for each risk factor in Eq. (7), and we calculate the risk probability for drowning or drifting by combining the probabilities for all risk factors:

$$Risk = \alpha_{hour}x_1 + \alpha_{wind\ direction}x_2 + \ldots + \alpha_{fuzzy\ current}x_{x-1} + \alpha_{fuzzy\ wind}x_k = \sum_{n=1}^{k} \alpha_{risk\ factor}x_n \quad (10)$$

After performing risk assessment and prediction of drowning and drifting accidents, the results are collected to perform overall risk assessment for the beach; the computed values of the entire risk are then defined by selecting the value having the highest risk among the risks obtained from drowning and drifting accidents in the overall risk assessment. Equation (10) gives the method of obtaining the value. Parameter α is constant for each risk factor, and k is the number of risk factors. We used 13 features in the experiments.

4. Experiment results of Beach Risk Assessment

4.1 Data Set

We used the public data set provided by KMA [17] and KHOA [18]. The data set provided by KMA includes information, such as air temperature, rainfall, wind, humidity, and atmospheric pressure, which are provided by 510 stations including data from 1997 to the present, and can be collected in the units of minute, hour, day, month, and year. The data set provided by KHOA includes various data, such as tide, water height, current direction and current speed, water temperature, air temperature and atmospheric pressure, wind direction and wind speed, and other statistical data on ocean, and can be collected in the units of minute, hour, day, month and year.

Table 3 shows the public data used in this research, and risk assessment was performed each hour. When data was integrated on an hourly basis, the average value of the factor was used. In addition, research was performed for Daecheon beach having beach accident data from 2016–2018, and we collected 65 days data from Saturday of the third week of June to Sunday of the third week of August, which for Daechon beach is the open period.

Collected data is classified according to the beach risk factors analyzed in the previous study [6], and, as for the criteria for the classification, a total of 14 sub-risk factors including environmental factors and personal factors were defined through meetings with professionals in charge of safety at a beach, researchers, and related institutions. Also, drowning and drifting

accidents were defined to identify the type of accident, and a study was also performed to analyze the relations between the types of accidents and risk factors.

Table 3. Data sets used

Korea Meteorological Administration		
Name	**Type**	**Definition**
Time	Date	Month, day and time information
Wind direction	Number	Direction of wind
Wind speed	Number	Speed of wind

Korea Hydrographic and Oceanographic Agency		
Current direction	Number	Direction of current
Current speed	Number	Speed of current
Wave height	Number	Height of wave
Water temperature	Number	Water temperature information(°C)
Tide height	Number	Tide level

Table 4. High-level importance analysis results [6]

High-level	Weight(%)	Rank
Natural environment	**0.367**	**1**
Facility · Equipment· Manpower	0.259	2
Personal environment	0.207	3
Geographical environment	0.167	4

The results of the study show that the main factors of drowning accident are drinking, lack of skill of swimming, water height, heaving waves, rip current, and depth of water; and in the case of drifting accidents, the main factors are water height, heaving waves, rip current, drinking, and lack of skill of swimming. Also, the results of analysis on the relative importance for each of the main factors showed that environmental factors influence beach accidents more than personal factors, as revealed in **Table 4**.

4.2 Beach Risk Assessment

The hardware and operation system environments for the experiment are Intel Core i7-8700 3.20 GHz CPU, 16.0 GB RAM, Windows 10 64-bit and Python. In order to perform beach risk assessment, the study was performed with a previously defined data set, and experiments were performed with the data extracted from the Fuzzy Set and situation Awareness described in Section 3. In the experiments, risk assessment on drowning and drifting accidents and beach risk assessment were performed. In each experiment, the study was performed in consideration of the cases where only Situation Awareness was applied, where Fuzzy theory and Situation Awareness were applied together, and where feature correlation was applied. Fuzzy water height and depth of water (whdw), Fuzzy current, current speed and current direction, and Fuzzy wind, wind speed and wind direction were generated through Fuzzy theory.

Table 5. Results of feature correlation

Drift(%) / Drown(%)	Hour	Wind direction	Current direction	Wind speed	Current speed	Wave height	Water temp	Tide height	Dif tide height	Tide var	Fuzzy whdw	Fuzzy current	Fuzzy wind
Hour	-	0.20	0.04	0.08	-0.03	-0.01	0.06	0.19	0.02	0.01	0.04	-0.09	0.01
Wind direction	0.20	-	-0.09	0.21	-0.05	0.15	0.04	-0.06	0.04	0.02	0.12	0.03	0.26
Current direction	0.04	-0.09	-	-0.08	-0.22	-0.16	0.09	0.64	0.04	0.03	0.00	-0.65	-0.08
Wind speed	0.05	0.21	-0.08	-	-0.06	0.63	-0.07	-0.02	-0.02	-0.03	0.61	0.04	0.59
Current speed	-0.03	-0.05	-0.22	-0.06	-	-0.10	-0.02	0.05	-0.10	-0.12	-0.22	0.16	0.01
Wave height	-0.01	0.15	-0.16	0.63	-0.10	-	-0.14	-0.16	-0.03	-0.05	0.92	0.16	0.44
Water temp	0.06	0.04	0.09	-0.07	-0.02	-0.14	-	0.09	-0.16	-0.09	-0.09	-0.04	-0.12
Tide height	0.19	-0.06	0.64	-0.02	0.05	-0.16	0.09	-	-0.03	-0.04	0.07	-0.74	-0.06
Dif tide height	0.02	0.04	0.04	-0.02	-0.10	-0.03	-0.16	-0.03	-	0.74	0.02	-0.02	0.01
Tide var	0.01	0.02	0.03	-0.03	-0.12	-0.05	-0.09	-0.04	0.74	-	0.00	-0.01	0.01
Fuzzy whdw	0.04	0.12	0.00	0.61	-0.22	0.91	-0.10	0.07	0.02	0.00	-	0.01	0.39
Fuzzy current	-0.08	0.04	-0.65	0.04	0.17	0.16	-0.04	-0.73	-0.02	-0.01	0.01	-	0.05
Fuzzy wind	0.00	0.25	-0.09	0.58	0.01	0.46	-0.13	-0.06	0.01	0.02	0.40	0.05	-

Feature correlations were used to identify relations present among features in a data set, to which Pearson correlation coefficient was applied. **Table 5** shows the results of calculating correlations among features, which confirm that current direction and Fuzzy current, and tide height and Fuzzy current have more negative correlation in drowning than other features, and that wind speed and wave height, wind speed and Fuzzy whdw, wave height and Fuzzy whdw, and dif tide and tide variation have relatively more positive correlations.

In drifting, the results confirm that tide height and Fuzzy current have relatively high negative correlation, while wave height and wind speed, tide height and current direction, Fuzzy whdw and wind speed, and Fuzzy whdw and wave height have relatively high positive correlation, respectively. We combined two features having relatively high positive correlations to generate a new feature, while the features having relatively high negative correlations were not combined, and were proceeded with as is.

Table 6. Drowning and drifting risk assessment experiment results

Drifting accident			
Type	**TSS(%)**	**RSS(%)**	**R^2(%)**
Situation Awareness	19.8559	2.1585	0.8913
Situation Awareness + Fuzzy	19.8559	2.1294	0.8928
Situation Awareness + Fuzzy + Feature Correlation	19.8559	2.1568	0.8914
Drowning accident			
Type	**TSS(%)**	**RSS(%)**	**R^2(%)**
Situation Awareness	4.6303	0.6389	0.8621
Situation Awareness + Fuzzy	4.6303	0.6381	0.8622
Situation Awareness + Fuzzy + Feature Correlation	4.6303	0.6379	0.8623

Table 6 shows the results from the beach risk assessment experiment on drifting and drowning accidents, where K-fold cross validation was used for model validation, and K value was set to 10. We confirm the total sum of square (TSS), residual sum of square (RSS), Coefficient of Determination, and R2. R2 has value 0 to 1, and closer to value 1 represents better prediction. In the case of drifting accident, the prediction accuracy is shown to be 0.89 % on average, and in the case of drowning accident, the prediction accuracy is 0.86 % on average. Both the types of accident accidents are found to have less differences between the predictions of a model and the actual values through R2, and it can be confirmed that the proposed model can be used in beach risk assessment and prediction. From this fact, it can be seen that it is important to analyze focusing on the environment situations at the time.

Table 7. Beach risk assessment and prediction results

Drifting + drowning accident			
Type	**TSS(%)**	**RSS(%)**	**R^2(%)**
Situation Awareness	22.4295	2.0921	0.9067
Situation Awareness + Fuzzy	22.4295	1.9829	0.9116
Situation Awareness + Fuzzy+ Feature Correlation	22.4295	1.9843	0.9116

Then, study on the assessment and prediction of beach risk is conducted with the addition of drowning and drifting accident results. **Table 7** shows the results of experiment with beach

risk prediction accuracy of 0.90 % on average, and all the results of the 3 types of experiments are confirmed to show similar values.

We could find out that it is more efficient to assess and predict beach risks by combining drowning and drifting accidents; and that, in each case of drowning and drifting, there can be influences from relations between natural environment situations and other factors, but their influence is very low in the overall situation.

In conclusion, it is confirmed that the beach risk assessment model proposed in the experiment can be used in the assessment of actual beach risks, and that it can be used in predicting the beach risks assessment.

4.3 Impact of risk factors on beach accident

In order to confirm the natural environmental factors influencing beach risk for each hour, we experiment with the impact of beach risk factors on beach accident. **Table 8** shows the results of experiment, and the factor is selected that has the highest risk at the time the accident occurred, by comparison to the natural environmental factor data at the corresponding hour. Wind direction was the most influential factor, and the factors with similar influence include depth of water and water height. The next influential factors can be water temperature and tide height. Also, regarding water height, although it did not influence accident risk as a single factor, it can be seen that water height has high influence on accident by operating together with depth of water.

Table 8. Beach risk factors assessment results

Type	Total number	Type	Total number
Wind direction	94	Tide height	18
Current direction	7	Dif tide height	0
Wind speed	1	Tide variation	0
Current speed	0	Water height & depth (using Fuzzy)	93
Wave height	0	Current direction & speed (using Fuzzy)	0
Water temperature	22	Wind direction & speed (using Fuzzy)	0

5. Conclusion

Various factors are present at a beach, and each factor influences accidents occurring at the beach. Also, accidents are not caused by a single factor; various factors compositely operate in the occurrence. Research is needed to prevent beach risk accidents occurring from various factors in advance, and measuring and assessing risks of beach factors operating compositely. Therefore, this paper used natural environmental factors that can be extracted from public data to perform beach risk assessment, and preprocessing is conducted, such as noise removal, and normalization, since the measuring criteria are different for each factor in the corresponding data. For more precise risk assessment, Fuzzy theory and Situation Awareness are applied to the natural environmental factors extracted from public data, and the data set constructed this way is applied to Multi Linear Regression to measure and predict beach risks. Also, study was performed to identify the factors giving the largest influence on beach risk at a given time, and

to identify the correlations between each of the factors.

The results of experiment show that beach risks give a prediction accuracy of 0.90% on average, and the average prediction accuracy for drowning and drifting accident are 0.86% and 0.89%, respectively. These results confirmed that the beach risk assessment model proposed by this study could be utilized in assessing risks on a beach, and it is found, through feature correlation analysis, that the factors exchanging influences are different when drowning, as against drifting, accidents occur. These results mean that different natural environmental factors should be considered for each accident to predict and prevent drowning and drifting accidents. In addition, it was found, from the analysis of factors most influencing the accident that occurred at a beach at each time, that wind direction, depth of water, and water height showed the most influence. In particular, although water height never gave the highest influence on accidents as a single factor, it was found that the factor shows high influence by acting together with depth of water. From this fact, it is confirmed again that there exist correlations among factors on the occurrence of accidents, and that accidents occur due to the composite action of various factors.

In this paper, a model has been proposed for assessing the risk of a beach, and it can be confirmed from the results of experiment that risk assessment and prediction are available through the proposed model above a certain level. In addition, risk situation can be more precisely extracted by applying Situation Awareness for problems that have been known to be difficult for beach risk assessment in previous studies by using only natural environmental factors, and exact factor assessment becomes available by establishing the criteria on representative risk factors by using Fuzzy theory. In future studies, beach risk assessments will be conducted by adding personal and physical factors, in addition to natural environmental factors. For accurate beach risk analysis, assessment and prediction will be performed through improved Fuzzy theory and Situation Awareness. Also, we will study algorithms for improved beach risk assessment.

References

[1] Weir and Adam, "Border to border–assessing the risks of an entire coastline," *A Century of Lifesaving-a Challenge to Drowning Prevention*, p.170, 2013. Article (CrossRef Link).

[2] David Reid, "Nation Coastal Safety Report 2016 Injuries on Australian Beaches," *Surf Life Saving*, 2016. Article (CrossRef Link).

[3] Browne, Matthew, et al, "An intelligent system for remote monitoring and prediction of beach conditions," in *Proc. of the International Conference on Artificial Intelligence and Applications,* pp. 533-537, 2005. Article (CrossRef Link).

[4] Short, A. D., and C. L. Hogan, "Rip currents and beach hazards: their impact on public safety and implications for coastal management," *Journal of Coastal Research,* pp. 197-209, 1994.

[5] Korea Life Saving Society, "A Study on the Establishment of Efficient Safety management Standards for Beaches," 2014.

[6] Gun-Yoon Shin, Dong-Wook Kim, et al, "Beach Risk Assessment Based on Multiple linear Regression for uncertain management of Drifting and Drowning," *Journal of Korean Institute of Intelligent*, vol. 29, no. 1, pp. 9-15, February, 2019. Article (CrossRef Link).

[7] Fisher L and Lewis D, "RNLI Public Rescue Equipment (PRE) Recommendations Report Ventnor Beach, Ventnor Town Council," *lifeboats*, August, 2011. Article (CrossRef Link).

[8] NSW government, "Coastal Public Safety Risk Assessment Newcastle Local Government Area," June, 2016. Article (CrossRef Link).

[9] Byungmo Y, Jooyong L, et al, "Quantitative Risk Assessment for Beach Drowning Management," in *Proc. of The 3rd International Rip Current Symposium*, vol. 72, pp. 117-121, 2014. Article (CrossRef Link).

[10] Shapiro, Arnold F, "Fuzzy regression models," *Article of Penn State University*, vol. 102, no. 2, pp. 373-383, 2005.

[11] Song, Kyung-Bin, et al, "Short-term load forecasting for the holidays using Fuzzy linear regression method," *IEEE transactions on power systems*, vol. 20, no. 1, pp. 96-101, 2005. [Article (CrossRef Link)](#).

[12] Endsley, Mica R, "Toward a theory of situation awareness in dynamic systems," *Situational Awareness*, 99, 9-42, 2017. [Article (CrossRef Link)](#).

[13] Mitsch, Stefan, et al, "A survey on clustering techniques for situation awareness," in *Proc. of Asia-Pacific Web Conference*, pp. 815-826, 2013. [Article (CrossRef Link)](#).

[14] Haidar, Ali, and Brijesh Verma, "A genetic algorithm based feature selection approach for rainfall forecasting in sugarcane areas," in *Proc. of 2016 IEEE Symposium Series on Computational Intelligence*, pp. 1-8, 2016. [Article (CrossRef Link)](#).

[15] Liu, Yu, Qinghua Guo, and Maggi Kelly, "A framework of region-based spatial relations for non-overlapping features and its application in object based image analysis," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 63, no. 4, pp. 461-475, 2008. [Article (CrossRef Link)](#).

[16] Benesty, Jacob, Jingdong Chen, and Yiteng Huang, "On the importance of the Pearson correlation coefficient in noise reduction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 757-765, 2008. [Article (CrossRef Link)](#).

[17] Korea Meteorological Administration. [Article (CrossRef Link)](#).

[18] Korea Hydrographic and Oceanographic Agency. [Article (CrossRef Link)](#).

Gun-Yoon Shin received his Bachelor's degree in Interactive Media Convergence from Gachon University, Korea in 2017, and Master's degree in Computer Engineering from Gachon University, Korea in 2018. He is currently a Ph.D. candidate in the Department of Computer Engineering, Gachon University, Korea. His research interests include Authorship Attribution, Genetic Algorithm, Information Security, Data Mining, and Machine Learning

Sung-Sam Hong received his Master's degree in Computer Science from Gachon University, Korea in 2011, and Ph.D. degree in Computer Science from Gachon University, Korea in 2016. He is a researcher professor in the Department of Computer Engineering of Gachon University, Korea. His research interests include Multimedia Security, Intelligent Information Security, Machine Learning, Cryptology, Data Mining, and Big Data..

Dong-Wook Kim received his Bachelor's degree in Computer Software from Gachon University, Korea in 2015, and Master's degree in Computer Engineering from Gachon University, Korea in 2017. He is currently a Ph.D. candidate in the Department of Computer Engineering, Gachon University, Korea. His research interests include Insider Threat, Information Security, Data Mining, and Machine Learning.

Cheol-Hun Hwang received his Bachelor's degree in Computer Engineering from Gachon University in 2019, Korea. Currently, he is a Master's student in the Department of Software, Gachon University, Korea.

Myung-Mook Han received his M.S. degree in computer science from New York Institute of Technology in 1987, and Ph.D. in information engineering from Osaka City University in 1997, respectively. From 2004 to 2005, he was a visiting professor at Georgia Tech Information Security Center (GTISC), Georgia Institute of Technology. Currently, he is a professor in the Department of Software, Gachon University, Korea. His research interests include Information Security, Intelligent System, Data Mining, and Big Data.

Hwayoung Kim received his B.S. degree in Maritime Transportation from Mokpo National Maritime University in 1998, and M.S. degree in Engineering of Maritime Information System from Mokpo National Maritime University, Korea, in 2004. He received his Ph.D. in Engineering of Urban Environmental System from Kyushu University in Japan in 2007. Currently, he is working as assistant professor in the Division of Maritime Transportation in Mokpo National Maritime University, Korea. His research interests include data mining, evaluation of efficiency, environmental assessment and fuzzy system applied to maritime transportation, port logistics, and maritime safety.

Young jae Kim received his Master's of Business Administration from Soongsil University, Korea, in 2010. Currently, he is working as head of department at GIS SW Company. His research interests include GIS, big data, and artificial intelligence.