

LETTER • OPEN ACCESS

Prediction of coastal flooding risk under climate change impacts in South Korea using machine learning algorithms

To cite this article: Sang-Jin Park and Dong-Kun Lee 2020 *Environ. Res. Lett.* **15** 094052

View the [article online](#) for updates and enhancements.

Recent citations

- [Flooding trends and their impacts on coastal communities of Western Cape Province, South Africa](#)
Kaitano Dube *et al*

Environmental Research Letters



LETTER

OPEN ACCESS

RECEIVED

27 March 2020

REVISED

5 July 2020

ACCEPTED FOR PUBLICATION

14 July 2020

PUBLISHED

25 August 2020

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Prediction of coastal flooding risk under climate change impacts in South Korea using machine learning algorithms

Sang-Jin Park¹ and Dong-Kun Lee²

¹ Interdisciplinary Program in Landscape Architecture, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea

² College of Agriculture and Life Science, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea

E-mail: dklee7@snu.ac.kr

Keywords: coastal management, risk analysis, coastal vulnerability, natural hazards, climate change impacts, machine learning, regional climate models

Abstract

Coastal areas have been affected by hazards such as floods and storms due to the impact of climate change. As coastal systems continue to become more socially and environmentally complex, the damage these hazards cause is expected to increase and intensify. To reduce such negative impacts, vulnerable coastal areas and their associated risks must be identified and assessed. In this study, we assessed the flooding risk to coastal areas of South Korea using multiple machine learning algorithms. We predicted coastal areas with high flooding risks, as this aspect has not been adequately addressed in previous studies. We forecasted hazards under different representative concentration pathway climate change scenarios and regional climate models while considering ratios of sea level rise. Based on the results, a risk probability map was developed using a probability ranging from 0 to 1, where higher values of probability indicate areas at higher risk of compound events such as high tides and heavy rainfall. The accuracy of the average receiver operating characteristic curves was 0.946 using a k-Nearest Neighbor algorithm. The predicted risk probability in 10 year increments from the 2030s to the 2080s showed that the risk probability for southern coastal areas is higher than those of the eastern and western coastal areas. From this study, we determined that a probabilistic approach to analyzing the future risk of coastal flooding would be effective to support decision-making for integrated coastal zone management.

1. Introduction

Climate change is a severe threat to current and future generations. Natural hazards have become more unpredictable and occur more frequently and with greater force due to climate change (Berz *et al* 2001, Kundzewicz *et al* 2014). Coastal areas are threatened by hazards such as flooding, erosion, and storms (Klein *et al* 2003, Nicholls and Cazenave 2010, Saxena *et al* 2013, Lilai *et al* 2016) and will be more vulnerable in the future to predicted climate change impacts such as sea level rise and extreme weather events (Klein and Nicholls 1999, Lilai *et al* 2016, Voudoukas *et al* 2018). Furthermore, the number of people living in coastal areas globally is expected to increase from 1.8 to 5.2 billion by the 2080s (IPCC 2014, Bhable *et al* 2015, Neumann *et al* 2015). As coastal systems continue to become more socially and environmentally complex, the cost of damage from coastal hazards due to climate change impacts will also rise (Kleint *et al*

2001, IPCC 2007, Szlafsztein and Sterr 2007, Balica *et al* 2012).

Natural disasters in Korea are mostly caused by meteorological events (Yoon *et al* 2016, Azam *et al* 2017, Han *et al* 2018). The total damage caused by disasters in the last 10 years is attributed mainly to typhoons (49%) and heavy rain (40%) (Ministry of the Interior and Safety 2016). Since 1990, the numbers of heavy rain advisories and warnings have increased by 25% and 60%, respectively (Korea Meteorological Administration 2011). Additionally, South Korea is a peninsula with several large cities situated along the coast and 27.5% of its total population lives in coastal areas (Oh *et al* 2020). Thus, a series of hazard prevention plans and the identification of at-risk areas are crucial for these coastal areas (Tran *et al* 2008, Kourgialas and Karatzas 2011). In this study, the risk of coastal flood hazards was analyzed, focusing on flooding caused by extreme weather events and sea level rise.

Existing research on coastal flooding has primarily used quantitative indices to characterize risks. The index method expresses the vulnerability of an area using arithmetic operations that incorporate classified factors affecting hazards. Several studies have generated a coastal/composite vulnerability index for calculating and assessing of coastal hazards in different study areas by using different variables (Dwarakish *et al* 2009, Sankari *et al* 2015, Pantusa *et al* 2018, Sahana and Sajjad 2019). These index studies have focused primarily on analyzing vulnerabilities that may indicate the relative risks encountered by coastal areas, but do not specifically calculate the actual risks.

These previous studies also only analyzed current vulnerability or risk and did not analyze future risks; however, several others have analyzed future risks. They calculated the future risk of coastal hazards by estimating water level heights (Wahl *et al* 2016, Vousedoukas *et al* 2016). Unlike previous reports, this study quantitatively calculated coastal hazard risks and predicted future risks by considering the occurrence of compounding probabilistic events such as extreme precipitation and the rising tidal ratio. Although they analyzed the risk of coastal hazards using statistical and physically-based methods, there are some factors that could not be addressed. These studies did not obtain a spatial distribution of risk (Wahl *et al* 2016) and analyzed continental scales that are difficult to apply to regional scales (Vousedoukas *et al* 2016). Also, they did not consider uncertainty by comparing multiple models to predict future hazards. In addition, rainfall is a very important factor in flooding but they herein focus only on changes in water level.

In short, this study analyzed the actual risk probability, not a relative vulnerability, using a coastal flooding risk analysis that considers rainfall events as well as tidal levels, because the risk to coastal areas of heavy rainfall depends on the tide (Van Den Hurk *et al* 2015, Eilander *et al* 2020). Additionally, future coastal flooding risks were estimated by considering the actual rising rate of the tide and the forecasted rainfall according to different representative concentration pathway (RCP) climate change scenarios and regional climate models. Multiple machine learning (ML) algorithms that have been widely used in recent studies as part of probabilistic approaches were used to probabilistically calculate the coastal flood risk. The results of this study identify future at-risk areas and can support decision making for integrated coastal zone management (ICZM) in South Korea by identifying which areas require hazard prevention plans.

2. Materials and method

2.1. Study area

In this study, the spatial coverage is South Korea (33–38° N, 125–131° E). Summer in South Korea is generally hot and wet and typhoons that occur frequently

in July and August bring heavy rainfall to coastal areas. The heaviest rainfall in this region was recorded in mid-July and mid-August when the daily average rainfall ranged between 220 mm and 322 mm. The maximum daily rainfall recorded for the period of 1973–2010 was 870.5 mm in Gangneung, South Korea, on August 31, 2002 (Korea Meteorological Administration 2011).

The total length of the coastline of Korea is 14 962.8 km and the spatial scope of this research was set up 1 km from the coastline, following the ‘coastal management law’ in Korea. The coastline along the East Sea is monotonous, and the water depth is generally deep. The West Sea has a shallow coastline with an average depth of 44 m. The coastline of the South Sea is complex and contains numerous islands and harbors. Figure 1 shows the locations of the 68 weather observatories and 46 tide observatories in Korea.

2.2. ML algorithms

A coastal hazard risk analysis was implemented using three ML algorithms: k-nearest neighbor (kNN), random forest (RF), and support vector machine (SVM). Previous studies have frequently compared these three machine learning techniques in their prediction methods (Harefa and Pratiwi 2016, Potdar and Kinnerkar 2016; López-Serrano *et al* 2016, Danades *et al* 2017, Thanh Noi and Kappas 2017). The results of these algorithms were subsequently compared.

The kNN, proposed by Cover and Hart (1967), is an easy-to-implement supervised ML algorithm that is as simple as the Naive-Bayes Classifier (Jadhav and Channe 2016). The proximity of data points to one another affects the results of the algorithm (Bhavsar and Ganatra 2012, Kim *et al* 2012). In this study, the analysis was performed by setting *k* as 5, 10, and 15. The analysis was the best when *k* was set to 5. The RF algorithm is an ensemble learning method operated by constructing multiple decision trees during a training period, and it is frequently used in research along with ML techniques such as SVM and neural networks. More details related to RF are described by Breiman (2001). Setting the number of trees and depths is important; herein, these values were set to 10 and 3, respectively. The SVM algorithm, proposed by Cortes and Vapnik (1995), is a versatile ML algorithm in that it can be classified in unlabeled datasets. When dividing two sets to create the classifier hyperplane (Rebentrost *et al* 2014), various classifiers such as linear and non-linear forms can be generated according to the characteristics of the data. Therefore, this algorithm is a high-performing technique used to analyze real-world data (Tong and Koller 2001). Setting the kernel function is important in SVM since it helps to overcome shortcomings related to linear separability (López-Serrano *et al* 2016). In this study, we used the radial basis function (RBF) kernel function, which makes non-linear classifiers.

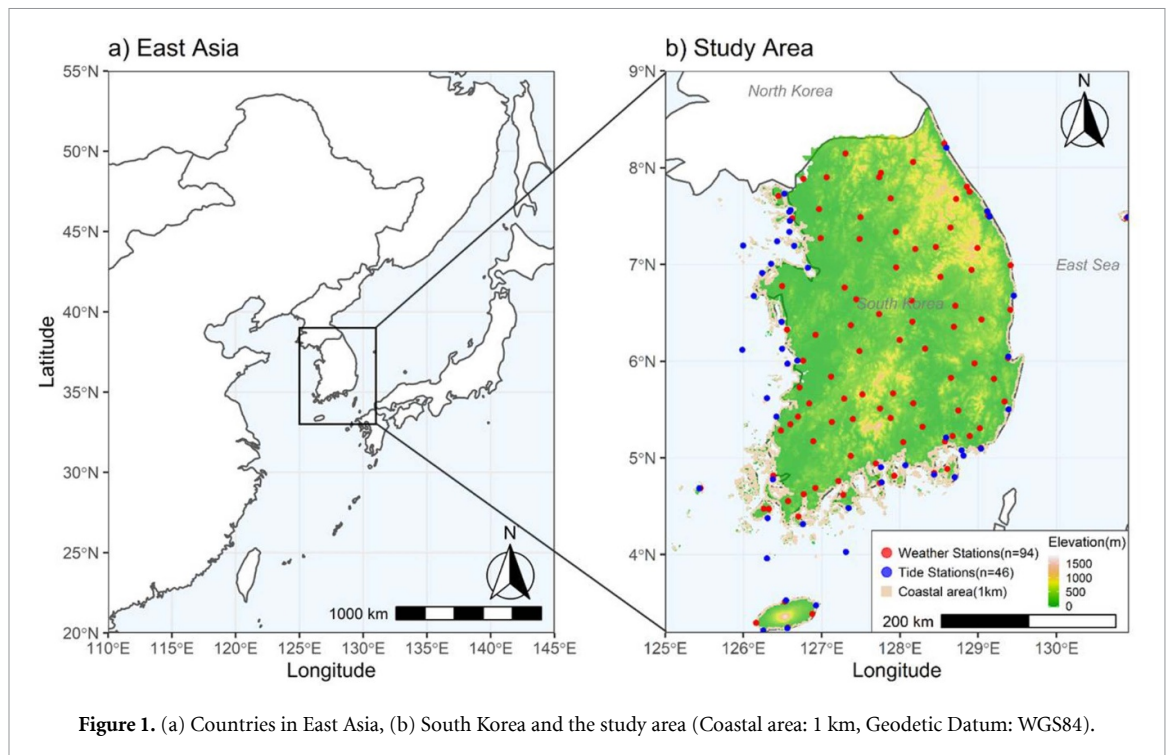


Figure 1. (a) Countries in East Asia, (b) South Korea and the study area (Coastal area: 1 km, Geodetic Datum: WGS84).

These ML algorithms were used in this study to compensate for the shortfalls of each individual algorithm (Hao *et al* 2019). Additionally, by using ML algorithms, we could consider complex and diverse influencing factors caused by climate change. The results of the algorithms were compared through approximately 700 iterations to reduce the uncertainty of the model itself.

2.3. Method

2.3.1. Data

The variables used in the analysis (tide, rainfall, elevation, slope, urban area, and grassland) were selected based on previous literature reviews (Mahendra *et al* 2010; Sankari *et al* 2015; Ashraful Islam *et al* 2016; Giardino *et al* 2018; Pantusa *et al* 2018), while rainfall, which was unaddressed in previous studies, was used to consider compounding events in this study. All data obtained for the risk analysis were transformed into a 1 km² grid because the raw data consisted of different points and polygons. All of the data were obtained from Official Korean Government websites (table 1). They were organized into a data table by day and grid. In addition, a map showing data for coastal flooding traces was obtained and used as the labeling data for classifying a machine learning algorithm. The grids on the map where coastal flooding occurred were labeled '1', or '0' if no flooding occurred.

2.3.2. Coastal risk analysis

The entire data table, in which all variables were organized by day and grid, was undersampled for the risk analysis due to imbalanced data that included

more cases of non-flooding than flooding. By undersampling the data, the risk analysis could be performed (He and Garcia 2009). The undersampled data were subsequently split into training (70%) and test (30%) datasets, after which the risk analysis was implemented using kNN, RF, and SVM. The procedure from the undersampling to running the algorithms was repeated 1200 times. After running the three ML algorithms, the results were compared using the receiver operating characteristic (ROC) accuracy scores and curves. ROC curves are mainly used to assess model accuracy, and the model is judged by the relationship between the false positive rate (1-Specificity) and the true positive rate. The risk probability was calculated, and risk maps were constructed using the results obtained from the three ML algorithms with the highest accuracy scores.

2.3.3. Prediction

The future risk probability was predicted using the highest performance algorithm. To predict future risks under the impacts of climate change, the continuous variables (rainfall and tidal level) were forecasted daily into the future. These were used to evaluate future coastal flooding risks.

Initially, the rainfall data used was the RCP AR5 scenario (4.5, 8.5) precipitation data produced from five different regional climate models (RCMs): CCLM, HadGEM3-RA, RegCM4, SNU-RCM, and WRF. These models were produced by the Regional Climate Detailing Project in East Asia (CORDEX-EA: Coordinated Regional Downscaling Experiment—East Asia, source: <http://cordex-ea.climate.go.kr/cordex>). These were obtained from

Table 1. List of variables.

Data (abbreviation)	Source (abbreviation, Website)	Type	Period
Mean tide (T)	Korea Hydrographic and Oceanographic Agency (KHOA, www.khoa.go.kr/eng/)	Point	2002–2014
Daily maximum rainfall (R)	Korea Meteorological Administration (KMA, https://web.kma.go.kr/eng/index.jsp)	Grid	2010
Elevation (E)	Ministry of Environment		
Slope (S)	(ME, https://eng.me.go.kr/eng/web/main.do)		
Urban area (U)			
Grassland (G)			
Coastal flood trace (CF)	Korea Land and Geospatial Informatrix Corporation (LX, www.lx.or.kr/eng.do)	Polygon	2002–2014
RCP precipitation data	KMA Climate Information Portal (KMA, http://climate.go.kr)	Ascii	2026–2085

the KMA climate information Portal (table 1), and KMA uses the HadGEM3-RA as its principal data. Daily maximum rainfall amounts for the different RCP scenarios of the RCMs were used, and monthly average rainfall values were calculated. Next, the tide was forecasted using real tidal range data obtained from a number of tidal observation stations. A Bayesian-influenced generalized additive model (GAM) was used to accurately forecast the tidal data as a sine-shaped curve with repetitive rising and falling trends. The Bayesian-influenced GAM, based on Bayes' theorem, performs regression while keeping functions smooth as a non-linear regression (Wood 2020). The past tidal pattern for each tidal station was analyzed and future tidal values were calculated and organized by day. The monthly average tidal values were then calculated from the daily tidal values.

Lastly, the calculated future tidal and rainfall data were then used to predict future risk probabilities. Future target periods ranged from the 2030s to the 2080s, with one division spanning 5 years (e.g. 2050: January 1, 2046–December 31, 2055). The process of running the model for prediction consisted of three steps (figure 2). (1) The kNN classifier was created similar to the manner in which it was created in the coastal risk analysis, (2) tidal and rainfall variables were each replaced by the predicted values, and (3) the replaced data table was used for prediction. This routine was repeated almost 1200 times to reduce uncertainty.

3. Results

3.1. Comparison of ML algorithms

As a result of running the three ML algorithms for the coastal risk analysis, the accuracy of the resulting average ROC curves of each algorithm were kNN (0.946), RF (0.938), and SVM (0.940), as shown in figure 3(a). The ROC curves of the other models did not appear

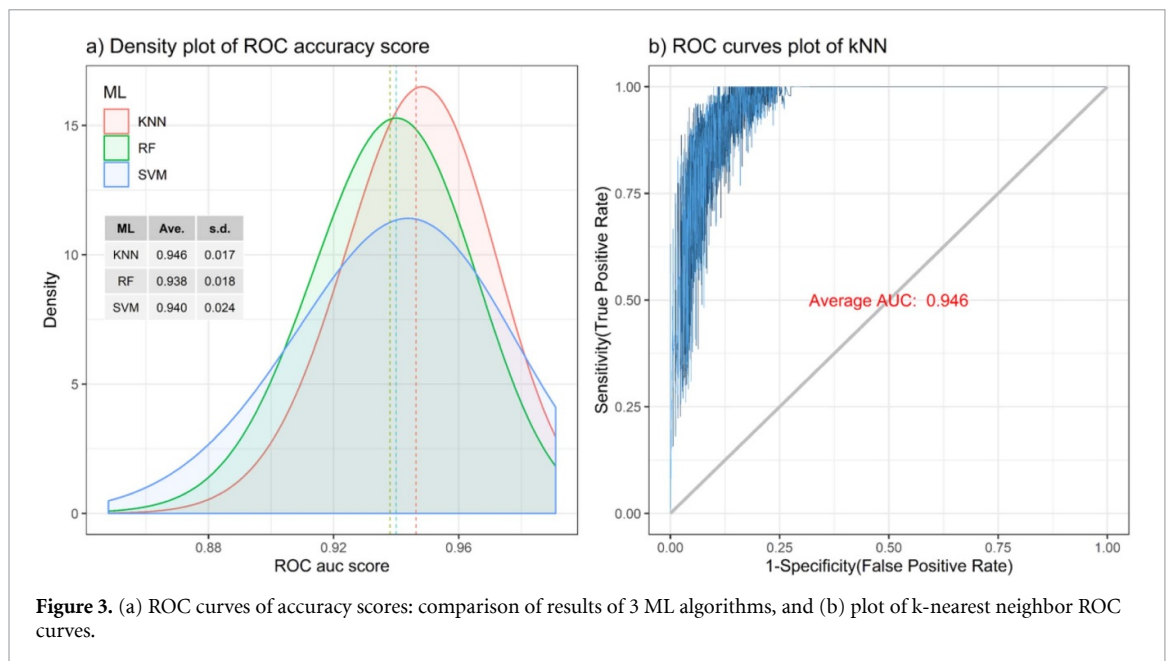
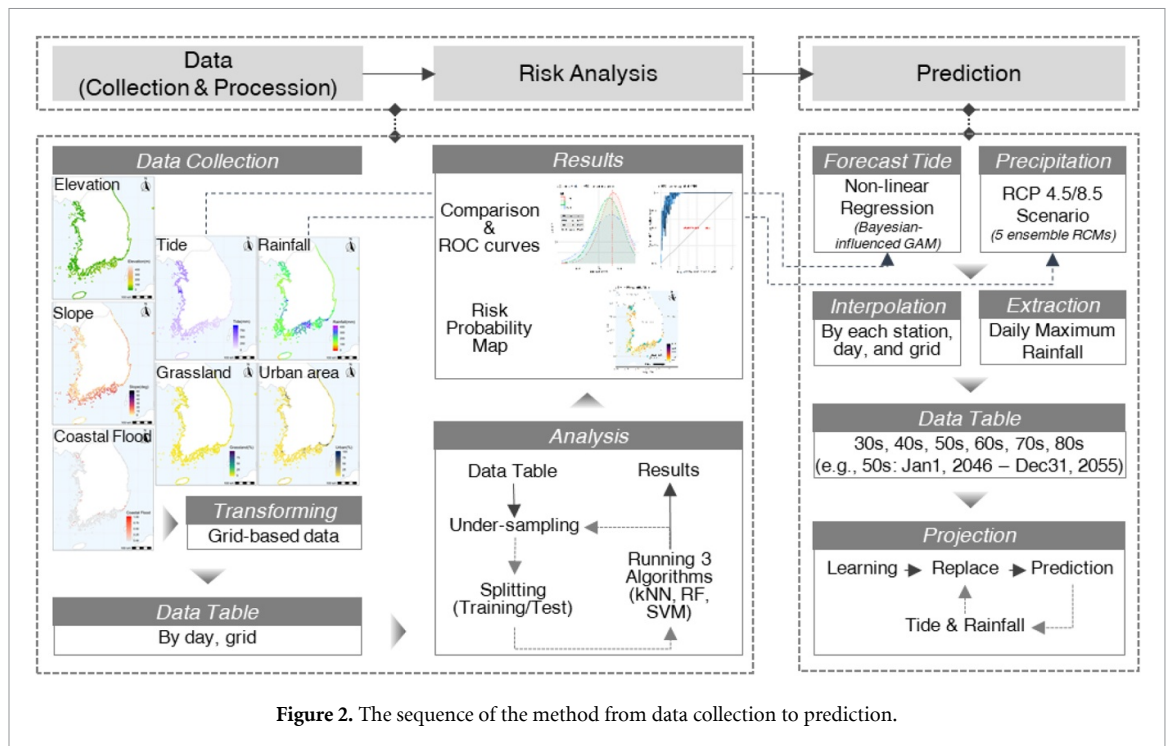
to be low either; however, the kNN model produced relatively better results. Also, figure 3(a) shows that the kurtosis at the density of the kNN ROC accuracy is slightly higher than the others, which means that the accuracy of the kNN is not as biased compared to the others. Therefore, the kNN was used for the final risk probability mapping and future prediction analysis. Figure 3(b) shows the trade-off between the false positive rate (1-Specificity) and the true positive rate. When the curves are closer to the upper left corner, its classifier exhibits good performance.

3.2. Risk probability map

A risk probability map with a gradational color distribution was developed, as shown in figure 4(a), based on the results of the model. The higher probabilities indicate areas at a higher risk of compounding events such as high tides and heavy rainfall. The blue dot on the graph indicates where coastal flooding occurred from 2002–2014. Comparing where the actual flooding occurred and was estimated to occur, the risk probability was relatively high in the area where the actual flooding occurred. Figure 4(b) compares the frequency in percentage by each class between the actual coastal flooding point mentioned above and the risk probability calculated by the kNN. According to this result, the value calculated by the kNN model overestimated the risk class below 0.75; however, it was approximately 64.35% accurate in estimating risk probabilities above 0.5. Even if the model for calculating risk probabilities tends to overestimate the risk, areas where risk probabilities of 0.5 or higher have been derived could still be at risk in the near future.

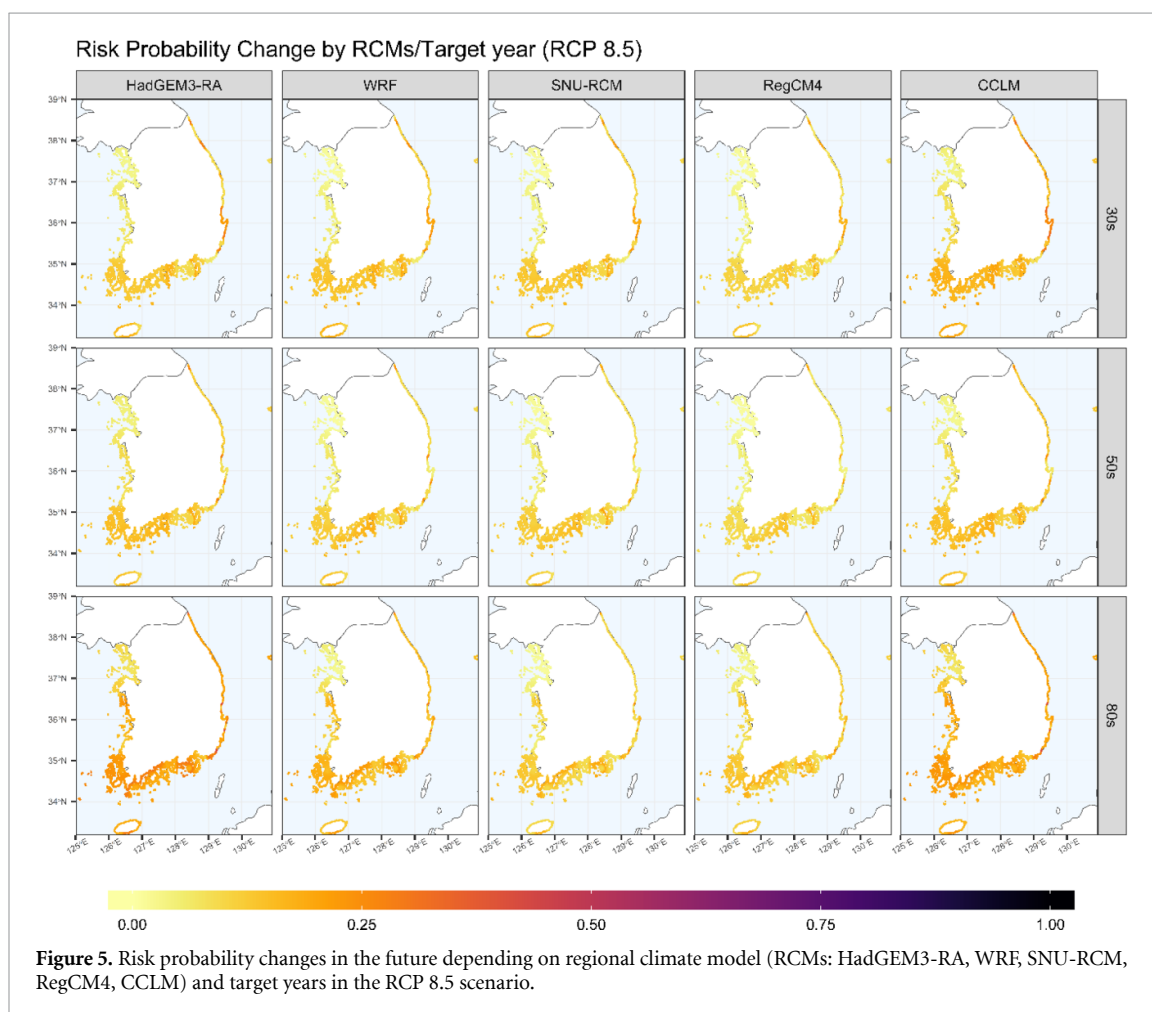
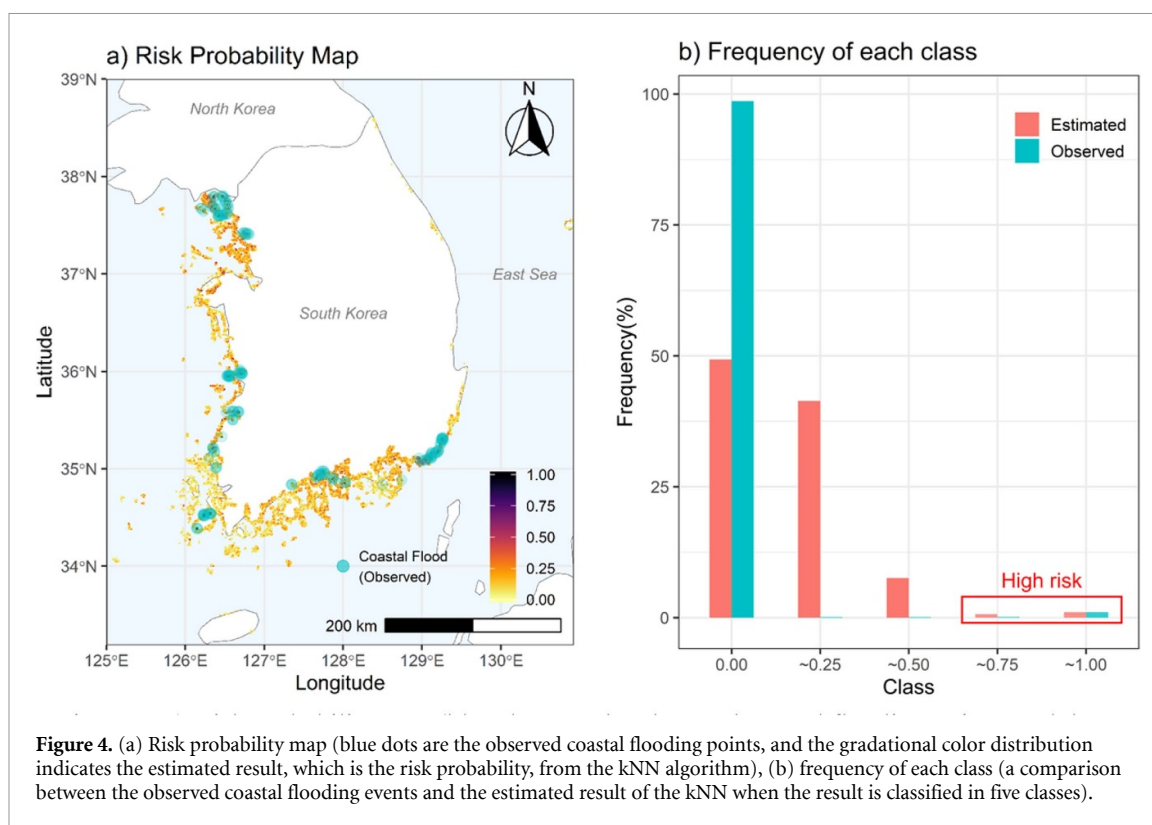
3.3. Future risk under climate change impacts

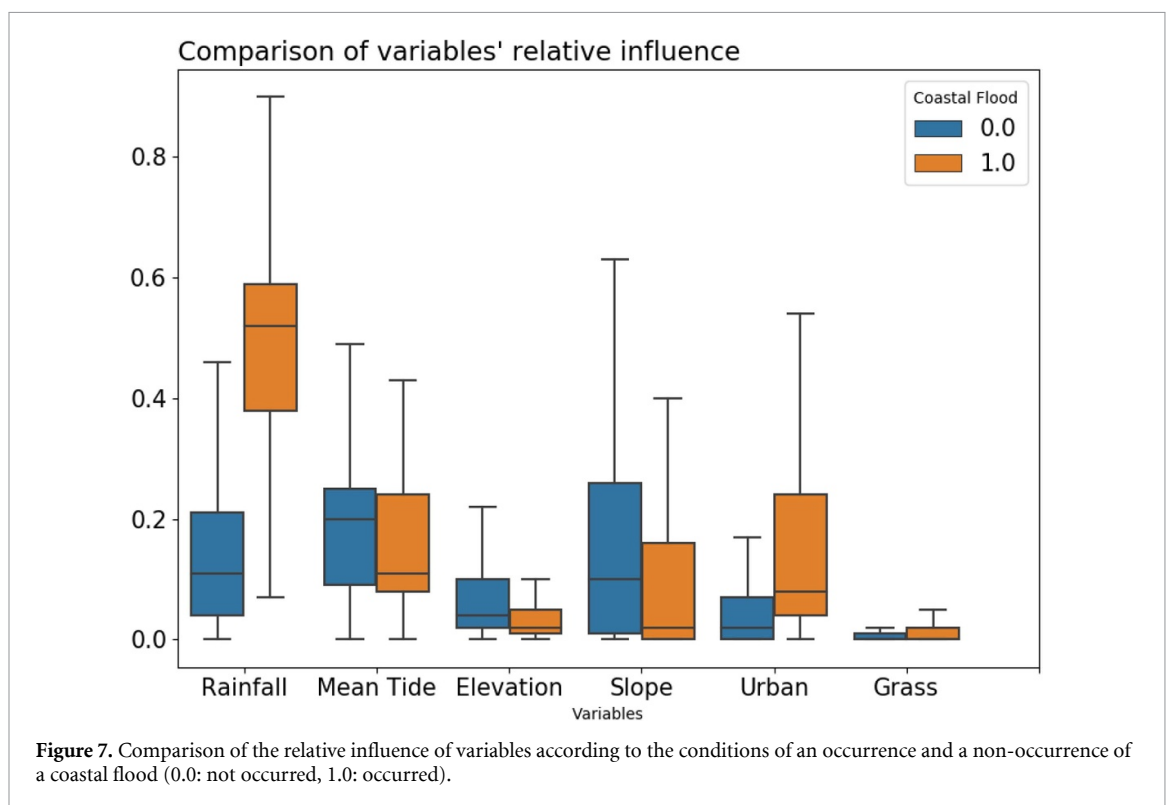
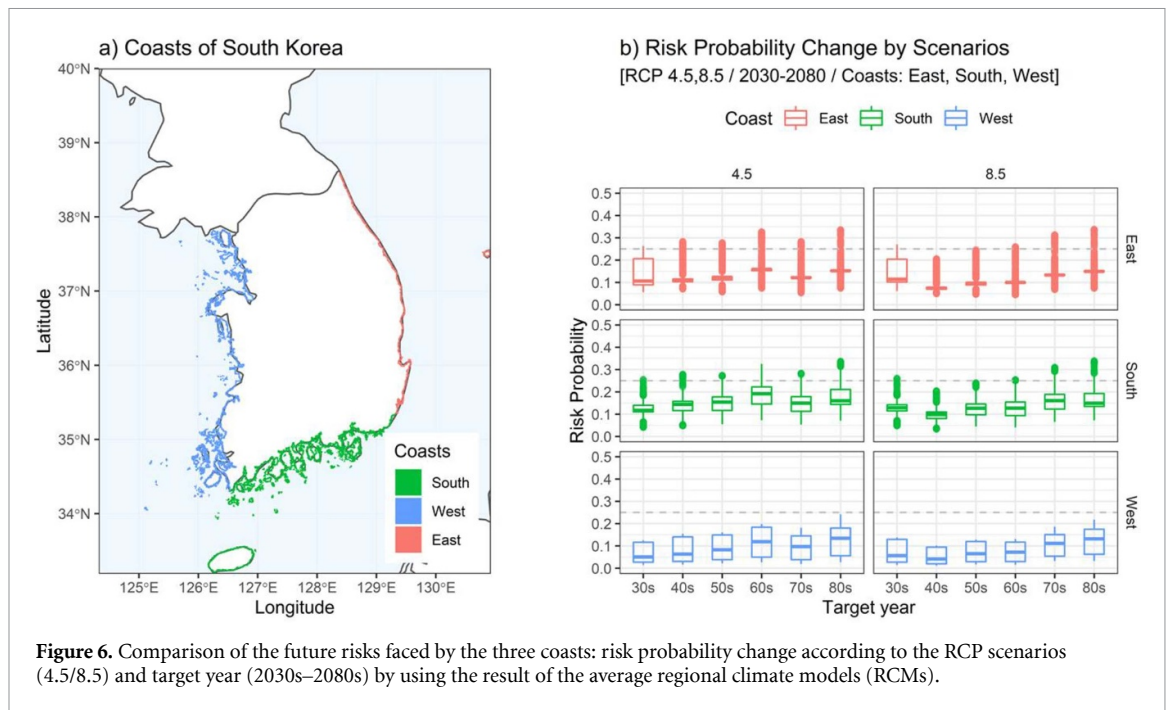
The prediction was implemented using monthly average rainfall and tidal values as described. In the process, rainfall predicted to occur in the future was used



as a density function to consider the uncertainty of future rainfall. Tidal data was input into the kNN classifier by month, the rainfall value was estimated by substituting the kernel density in one month, and the model was used to calculate the monthly predicted risk. Then, according to the comparison of the monthly risk probability data, the risk probability increased for the months of June, July, and August for most of the RCP 4.5 and 8.5 scenarios. Based on these data, the average risk probability for each scenario (RCP 4.5/8.5; the 2030s to the 2080s) in June, July, and August was calculated to create a maximum risk

probability map for the scenario. Figure 5 shows the future risk probability changes in the 2030s, 2050s, and 2080s for RCP 8.5, according to the five RCMs. In general, risk increases from the 2030s to the 2080s among the five RCMs, and in particular, from the 2050s to the 2080s for CCLM and HadGEM3-RA. Although there were differences among the results obtained using the five RCMs, the southern coastal areas will generally be more vulnerable than the eastern and western areas. The reason for these regional differences should be determined, as this could be significant for coastal zone management.





4. Discussion

4.1. Regional differences

The description of the impacts under different climate change scenarios in the Intergovernmental Panel on Climate Change (IPCC) report states that, as climate change progresses, the world will be affected differently by region (IPCC 2007, 2014). In Korea, which is surrounded by three seas, the geographical characteristics along the eastern, western, and

southern coasts are different (figure 1). Therefore, the impacts of climate change are expected to differ, and therefore the risks associated with coastal inundation also differ.

Figure 6 displays the change in risk probability for the three seas around South Korea from the present to the future. In these graphs, the southern coast shows slightly more risk than the other two coasts from the 2030s to the 2080s at both RCP scenarios (4.5, 8.5). The risk level also exhibits an increase in the 2060s

in the RCP 4.5 scenario, but the risk increases in the 2070s in the RCP 8.5 scenario. Although there is a difference in the time at which the risk increases, this suggests that the risk probabilities will increase in any scenario in the 2050s. Therefore, measures for long term adaptation (30 or 40 years from now) should be prepared.

4.2. Significance factor

The average tidal values for the three bordering seas (west, south, and east) are 546.56 mm, 222.16 mm, and 23.76 mm, respectively, in the 2050s. The average elevations are 22.8 m, 50.9 m, and 43.1 m, respectively. Theoretically, the western area should be the most vulnerable since the ratio of tidal rise is higher and the average elevation is lower than the others, as shown in figure 1. However, both the southern and western areas are also at risk, though the gap will gradually increase in the future. In order to determine the reason why the risk along the southern coast was estimated to be higher than that of the other two coasts, we investigated which variable dominated the results.

We divided the original data that was used for the risk analysis into whether the coastal flood event occurred or not. Then, each variable was normalized from 0 to 1 and the results were compared. The difference in rainfall between the normalized values according to floods that occurred was higher than the others. We inferred that rainfall is therefore a key factor compared to other variables, such as altitude or slope, in the risk analysis, as shown in figure 7 (Ward *et al* 2018). This also demonstrates that the water level does not fit the assumption that the western coastal area is theoretically more at risk, as shown in figures 5 and 6. Furthermore, the fact that urban areas are frequently flooded may suggest that coastal management plans such as building facilities for protection should account for the vulnerability of urban coastal areas.

4.3. Methodological implications

This study compared coastal flood risk analyses using three ML algorithms. As a result of the risk analyses, the results of the kNN analysis model exhibited the highest reliability and accuracy. This suggests a slightly different implication from that of other studies. In other studies that used various ML techniques, the accuracy of the algorithm was higher when using a ML technique such as artificial neural networks (ANN), SVM, or RF than those obtained when using kNN (Potdar and Kinnerkar 2016, López-Serrano *et al* 2016, Thanh Noi and Kappas 2017). We concluded that the performance of the kNN in this study was slightly higher because it may have been influenced by the difference in data quality. This study used traces of actual flooding for the risk analysis, and as a result of running the model with these data,

the accuracy was high. We infer that the risk analysis using kNN may be applied broadly as a quantitative technique, unlike studies with index methods, according to the spatial distribution of the flooded region data, regardless of region. In addition, the data-driven statistical method using ML algorithms as well as kNN is useful in terms of scalability, because it can account for various influences such as compounding events and can quickly adapt to the input of new data. Therefore, this quantitative approach could be effective for risk analysis.

Moreover, we attempted to consider uncertainty by using an ensemble method such as comparing the results from the five RCMs. The climate model itself does have uncertainty, and that uncertainty increases over time (Knutti and Sedláček 2013). Therefore, many studies consider uncertainty regarding future climate change using an ensemble approach that compares multiple models (Parker 2013). This could be the best way for decision-makers to communicate about future risks. In this study, trends were confirmed by comparing several climate models rather than one. We used a data-driven method instead of a model-driven method because it is difficult to confirm the uncertainty of future risk in a model-driven approach when using large amounts of data from the regional climate models used in the study.

5. Conclusions

Six variables were used to evaluate the future probability of coastal flooding events based on three different ML algorithms, namely kNN, RF, and SVM. All the data obtained for the model, such as tidal, rainfall, and elevation data, were converted into data in a 1 km² grid since each raw data type consisted of different points or polygons. Using the three ML classifiers method, the risk probability was calculated and the results of the ROC curves and their accuracy scores were compared. The average accuracy score of the kNN was the highest (0.946), and a risk probability map was developed using the results estimated by the kNN classifier. To evaluate future coastal flood risks due to climate change, tidal and rainfall data were used as a continuous value in prediction. For the RCP (4.5/8.5), daily maximum rainfall data for different RCMs from the 2030s to the 2080s (e.g. 2050s, 01.01.2046–12.31.2055) were used for model prediction and the kernel density was used as the input data for the prediction. In terms of the tidal level, the rising values of future tides were calculated by considering the rate of increase at each tidal station and forecasted using a Bayesian-influenced GAM. We estimated the future risk probability using forecasted tidal and future rainfall. As a result, the risk probability increased over time and the risk probability increased in the southern coastal areas more so than in the eastern or western coastal areas.

In this study, we argue that there are significant implications. We initially found that the results of the kNN were performed slightly better than the other methods by comparing three ML algorithms. This can be attributed to the characteristics of the kNN, according to the quality of the original data. It also infers that risk analysis using a simple ML algorithm such as kNN could be applied widely, regardless of region. Next, rainfall was identified as a significant factor in this study. This means that the possibility of flooding can be increased due to the uncertainty in forecasting future rainfall patterns due to climate change. Lastly, future coastal flood risk analysis was analyzed using ensemble methods from different RCMs. We considered future uncertainty, though it might be helpful for decision-makers to communicate about future risks by providing variances and trends from different results.

As in the previous claim, the ML technique used in this analysis exerts a powerful force when reliable data is used, but the results are not as sophisticated or deterministic as the results of a model-driven analysis, such as a hydrodynamic model. As long as there is uncertainty regarding climate change, a data-driven approach using ML may be easy for predictive analyses. Therefore, future studies could address the idea that the future tide and surge heights are calculated using a hydrodynamic model together, as in the work of Hoch *et al* (2019) and Muis *et al* (2020) to improve the quality of the results. This could also be aligned with precipitation or rainfall to consider compound flooding of discharge-tidal interactions through statistical analysis, as in the work of Eilander *et al* (2020). In addition, shoreline changes could be included in the analysis of coastal hazard risks, making the results more meaningful for ICZM. Furthermore, for the purposes of predicting future risk, it was generally assumed that variables other than tide and rainfall would not change over time. Geographical factors such as elevation and slope might not vary with time, but land cover such as urban areas and grasslands will change over time. Thus, according to land cover changes, the future spatial distribution of risk probability could be different. However, the rainfall was a key factor in this analysis as described previously, so the land cover change might not affect strongly the prediction. Therefore, it would be informative if a similar study could be conducted that accounts for social and economic changes in the risk analysis.

Acknowledgments

This work was supported by the Korea Environmental Industry and Technology Institute (KEITI) grant funded by the Korea government (ME) (No. 2014001310007).

Data availability

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iD

Sang-Jin Park  <https://orcid.org/0000-0003-4240-5682>

References

- Ashraful Islam M, Mitra D, Dewan A and Akhter S H 2016 Coastal multi-hazard vulnerability assessment along the Ganges deltaic coast of Bangladesh—a geospatial approach *Ocean Coast. Manage.* **127** 1–15
- Azam M, Kim H S and Maeng S J 2017 Development of flood alert application in Mushim stream watershed Korea *Int. J. Disaster Risk Reduct.* **21** 11–26
- Balica S F, Wright N G and Van der Meulen F 2012 A flood vulnerability index for coastal cities and its use in assessing climate change impacts *Nat. Hazards* **64** 73–105
- Berz G, Kron W, Loster T, Rauch E, Schimetschek J, Schmieder J, Siebert A, Smolka A and Wirtz A 2001 World map of natural hazards - a global view of the distribution and intensity of significant exposures *Nat. Hazards* **23** 443–65
- Bhable S, Kayte S, Mali S, Kayte J N and Maher R 2015 A review paper on coastal hazard *J. Eng. Res. Appl.* **5** 83–93 (www.ijera.com/papers/Vol5_issue12/Part%20-%202/L512028393.pdf)
- Bhavsar H and Ganatra A 2012 A comparative study of training algorithms for supervised machine learning *Int. J. Soft Comput. Eng.* **2** 74–81
- Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32
- Cortes C and Vapnik V 1995 Support-vector networks *Mach. Learn.* **20** 273–97
- Cover T and Hart P 1967 Nearest neighbor pattern classification *IEEE Trans. Inf. Theory* **13** 21–7
- Danades A, Pratama D, Anggraini D and Anggraini D 2017 Comparison of accuracy level K-Nearest Neighbor algorithm and support vector machine algorithm in classification water quality status 2016 6th *Int. Conf. on System Engineering and Technology (ICSET)* pp 137–41
- Dwarakish G S, Vinay S A, Natesan U, Asano T, Kakinuma T, Venkataramana K, Pai B J and Babita M K 2009 Coastal vulnerability assessment of the future sea level rise in Udupi coastal zone of Karnataka state, west coast of India *Ocean Coast. Manage.* **52** 467–78
- Eilander D, Couasnon A, Ikeuchi H, Muis S, Yamazaki D, Winsemius H and Ward P J 2020 The effect of surge on riverine flood hazard and impact in deltas globally *Environ. Res. Lett.* accepted (<https://doi.org/10.1088/1748-9326/ab8ca6>)
- Giardino A, Nederhoff K and Voudoukas M 2018 Coastal hazard risk assessment for small islands: assessing the impact of climate change and disaster reduction measures on Ebeye (Marshall Islands) *Reg. Environ. Change* **18** 2237–48
- Han H J, Kim J H, Chung S E, Park J H and Cheong H K 2018 Estimation of the national burden of disease and vulnerable population associated with natural disasters in Korea: heavy precipitation and typhoon *J. Korean Med. Sci.* **33** 1–3
- Hao T, Elith J, Guillera-Aroita G and Lahoz-Monfort J J 2019 A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD *Divers. Distrib.* **25** 839–52
- Harefa J, Alexander A and Pratiwi M 2016 Comparison classifier: support vector machine (SVM) and K-Nearest Neighbor

- (K-NN) in digital mammogram images *Juisi* **02** 35–40 (<http://peipa.essex.ac.uk/pix/mias/>)
- He H and Garcia E A 2009 Learning from imbalanced data *IEEE Trans. Knowl. Data Eng.* **21** 1263–84
- Hoch J M, Eilander D, Ikeuchi H, Baart F and Winsemius H C 2019 Evaluating the impact of model complexity on flood wave propagation and inundation extent with a hydrologic-hydrodynamic model coupling framework *Nat. Hazards Earth Syst. Sci.* **19** 1723–35
- IPCC 2007 Climate change, impacts, adaptation and vulnerability *Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, ed M L Parry, O F Canziani, J P Palutikof, P J van der Linden and C E Hanson (New York: Cambridge University Press)
- IPCC 2014 Climate Change 2014 Part A: global and sectoral aspects Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (www.ipcc.ch/site/assets/uploads/2018/02/WGIIAR5-PartA_FINAL.pdf)
- Jadhav S D and Channe H P 2016 Comparative study of K-NN, Naive Bayes and Decision Tree Classification techniques *Int. J. Sci. Res.* **5** 1842–5
- Kim J, Kim B-S and Savarese S 2012 Comparing image classification methods: K-Nearest-Neighbor and support-vector-machines *Applied Mathematics in Electrical and Computer Engineering* pp 133–8
- Klein R J T and Nicholls R J 1999 Assessment of coastal vulnerability to climate change *Ambio* **28** 182–7
- Klein R J T, Nicholls R J and Thomalla F 2003 Resilience to natural hazards: how useful is this concept? *Environ. Hazards* **5** 35–45
- Kleint R J T, Nicholls R J, Ragoonaden S, Capobianco M, Aston J and Buckley E N 2001 Technological options for adaptation to climate change in coastal zones *J. Coast. Res.* **17** 531–43
- Knutti R and Sedláček J 2013 Robustness and uncertainties in the new CMIP5 climate model projections *Nat. Clim. Change* **3** 369–73
- Korea Meteorological Administration 2011 Korea Climate Change White Paper pp 1–117
- Kourgialas N N and Karatzas G P 2011 Gestion des inondations et méthode de modélisation sous SIG pour évaluer les zones d'aléa inondation-une étude de cas *Hydrol. Sci. J.* **56** 212–25
- Kundzewicz Z W *et al* 2014 Flood risk and climate change: global and regional perspectives *Hydrol. Sci. J.* **59** 1–28
- Lilai X, Yuanrong H, Wei H and Shenghui C 2016 A multi-dimensional integrated approach to assess flood risks on a coastal city, induced by sea-level rise and storm tides *Environ. Res. Lett.* **11** 014001
- López-Serrano P M, López-Sánchez C A, Álvarez-González J G and García-Gutiérrez J 2016 A comparison of ML techniques applied to Landsat-5 TM spectral data for biomass estimation *Can. J. Remote Sens.* **42** 690–705
- Mahendra R S, Mohanty P C, Srinivasa Kumar T, Shenoi S S C and Nayak S R 2010 Coastal multi-hazard vulnerability mapping: a case study along the coast of Nellore District, East Coast of India *Ital. J. Remote Sens.* **42** 67–76
- Ministry of the Interior and Safety 2016 Statistical Yearbook of Natural Disaster
- Muis S, Apecechea M I, Dullaart J, de Lima Rego J, Madsen K S, Su J, Yan K and Verlaan M 2020 A high-resolution global dataset of extreme sea levels, tides, and storm surges, including future projections *Front. Mar. Sci.* **7** 1–15
- Neumann B, Vafeidis A T, Zimmermann J and Nicholls R J 2015 Future coastal population growth and exposure to sea-level rise and coastal flooding - a global assessment *PLoS ONE* **10** 3
- Nicholls R J and Cazenave A 2010 Sea-level rise and its impact on coastal zones *Science* **328** 1517–20
- Oh H M *et al* 2020 Wave risk assessment on coastal areas in Korea APAC 2019, ed N Trung Viet, D Xiping and T Thanh Tung (Singapore: Springer) pp 1351–8
- Pantusa D, D'Alessandro F, Riefole L, Principato F and Tomasicchio G R 2018 Application of a coastal vulnerability index. A case study along the Apulian Coastline, Italy *Water (Switzerland)* **10** 1–16
- Parker W S 2013 Ensemble modeling, uncertainty and robust predictions *Wiley Interdiscip. Rev. Clim. Change* **4** 213–23
- Potdar K and Kinnerkar R 2016 A comparative study of ML algorithms applied to predictive breast cancer data *Int. J. Sci. Res.* **5** 1550–3
- Rebentrost P, Mohseni M and Lloyd S 2014 Quantum support vector machine for big data classification *Phys. Rev. Lett.* **113** 1–5
- Sahana M and Sajjad H 2019 Vulnerability to storm surge flood using remote sensing and GIS techniques: a study on Sundarban Biosphere Reserve, India *Remote Sens. Appl. Soc. Environ.* **13** 106–20
- Sankari T S, Chandramouli A R, Gokul K, Surya S S M and Saravanavel J 2015 Coastal vulnerability mapping using geospatial technologies in Cuddalore-Pichavaram Coastal Tract, Tamil Nadu, India *Aquat. Procedia* **4** 412–8
- Saxena S, Geethalakshmi V and Lakshmanan A 2013 Development of habitation vulnerability assessment framework for coastal hazards: cuddalore coast in Tamil Nadu, India-a case study *Weather Clim. Extremes* **2** 48–57
- Szlafsztein C and Sterr H 2007 A GIS-based vulnerability assessment of coastal natural hazards, state of Pará, Brazil *J. Coast. Conserv.* **11** 53–66
- Thanh Noi P and Kappas M 2017 Comparison of random forest, k-Nearest Neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 Imagery *Sensors* **18** 1
- Tong S and Koller D 2001 Support vector machine active learning with applications to text classification *J. Mach. Learn. Res.* **2** 45–66
- Tran P, Shaw R, Chantry G and Norton J 2008 GIS and local knowledge in disaster management: a case study of flood risk mapping in Thua Thien Hue province, Vietnam *Disasters* **33** 152–69
- Van Den Hurk B, Van Meijgaard E, De Valk P, Van Heeringen K J and Gooijer J 2015 Analysis of a compounding surge and precipitation event in the Netherlands *Environ. Res. Lett.* **10** 035001
- Vousdoukas M I, Mentaschi L, Voukouvalas E, Verlaan M, Jevrejeva S, Jackson L P and Feyen L 2018 Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard *Nat. Commun.* **9** 1–12
- Vousdoukas M I, Voukouvalas E, Mentaschi L, Dottori F and Giardino A 2016 Developments in large-scale coastal flood hazard mapping *Nat. Hazards Earth Syst. Sci.* **16** 1841–53
- Wahl T, Plant N G and Long J W 2016 Probabilistic assessment of erosion and flooding risk in the northern Gulf of Mexico *J. Geophys. Res.: Oceans* **121** 3029–43
- Ward P J, Couasnon A, Eilander D, Haigh I D, Hendry A, Muis S, Veldkamp T I E, Winsemius H C and Wahl T 2018 Dependence between high sea-level and high river discharge increases flood hazard in global deltas and estuaries *Environ. Res. Lett.* **13** 8
- Wood S N 2020 Inference and computation with generalized additive models and their extensions *Test* **29** 307–39
- Yoon D K, Kang J E and Brody S D 2016 A measurement of community disaster resilience in Korea *J. Environ. Plan. Manage.* **59** 436–60