

Application Programming Interface for Flood Forecasting from Geospatial Big Data and Crowdsourcing Data

<https://doi.org/10.3991/ijim.v13i11.11237>

Supattra Puttinaovarat ^(✉)
Prince of Songkla University, Surat Thani, Thailand
supattra.p@psu.ac.th

Paramate Horkaew
Suranaree University of Technology, Nakhon Ratchasima, Thailand

Abstract—Nowadays, natural disasters tend to increase and become more severe. They do affect life and belongings of great numbers of people. One kind of such disasters that happen frequently almost every year is floods in all regions across the world. A preparation measure to cope with upcoming floods is flood forecasting in each particular area in order to use acquired data for monitoring and warning to people and involved persons, resulting in the reduction of damage. With advanced computer technology and remote sensing technology, large amounts of applicable data from various sources are provided for flood forecasting. Current flood forecasting is done through computer processing by different techniques. The famous one is machine learning, of which the limitation is to acquire a large amount big data. The one currently used still requires manpower to download and record data, causing delays and failures in real-time flood forecasting. This research, therefore, proposed the development of an automatic big data downloading system from various sources through the development of application programming interface (API) for flood forecasting by machine learning. This research relied on 4 techniques, i.e., maximum likelihood classification (MLC), fuzzy logic, self-organization map (SOM), and artificial neural network with RBF Kernel. According to accuracy assessment of flood forecasting, the most accurate technique was MLC (99.2%), followed by fuzzy logic, SOM, and RBF (97.8%, 96.6%, and 83.3%), respectively.

Keywords—Application Programming Interface, Big Data, Flood Forecasting

1 Introduction

Floods are a kind of natural disasters that happen almost every year in each region across the world, with effects in wide areas [1]. In the past, it was unable to predict or forecast floods accurately, e.g., which areas and what time floods would happen; due to the shortage of data or insufficient data for flood forecasting [2]. Technological limitations were also taken into account, e.g., computer capacity for big data pro-

cessing, which required high-performance computers. Unfortunately, those computers were expensive and mostly used in large organizations [3]. These brought flood forecasting limitations. In addition, communication systems among organizations or agencies in charge of measurement, implementation, storage, and dissemination of involved data or factors influencing floods did not contain available channels. On the other hand, they still relied on traditional methods, e.g., sending data via file uploading. The limitation of those methods was that they still required manpower to send data to agencies or demanding persons. This brought delays and failures to send new data automatically [4], which finally led to obstacles to use data for flood forecasting. Insufficient or outdated data affected forecasting accuracy. However, with capabilities of modern technologies with fast and non-stop advancement, some can be used to solve the issues or limitations aforementioned. According to literature review relating to flood forecasting, there are currently 2 sciences of forecasting, i.e., hydrological theory and techniques [5], and machine learning theory [6]. The studies revealed various models of both theories for flood forecasting. In term of hydrology, the famous models for flood forecasting include Mike11 [7,8], Mike21 [9,10], HecRAS [11,12], and HecGeoRAS [13,14]. As for machine learning, the famous models or techniques for flood forecasting include artificial neural network (ANN) [15-17], fuzzy logic [18-20], decision tree [21-23], maximum likelihood [24-26], and self-organization map (SOM) [27-29]. The results of flood forecasting revealed close accuracy between hydrology and machine learning [30-32]. Even so, when considering their suitability for big data processing, it was found that machine learning techniques are widely used for forecasting in other aspects, i.e., stock price forecasting [33,34], health and medical analytics [35], weather forecasting, and forecasting other kinds of disasters [37-39]. When using big data and machine learning for flood forecasting, traditional data is still mostly relied on. To clarify, data acquired still requires manpower to search and to collect for forecasting. In some cases, data is not acquired in time; and thus unable to be used. Thus, this research proposed a method to reduce the limitations of traditional methods, through the development of automatic big data downloading system by application programming interface (API) from involved sources of data, i.e., Thailand Meteorological Department (TMD), GLOFAS [40], and database system. The purpose is to use data for flood forecasting by machine learning techniques, i.e., MLC, fuzzy logic, SOM, and RBF, which will bring accurate and in-time flood forecasting in order to use data further for flood monitoring and warning; and for efficient flood management.

2 Literature Review

According to the literature review, famous methods or techniques for flood forecasting include machine learning [15-29] and hydrological techniques [7-14]. However, most involved studies revealed the frequently used technique for forecasting works by big data, i.e., machine learning, rather than hydrological techniques [33-39] because hydrological models are mostly used with fixed factors that are less flexible to be increased as needed. As a result, forecasting accuracy is different in each particular

area due to different spatial contexts. The studies revealed further that using suitable factors for spatial contexts could increase forecasting accuracy. This research, therefore, relied on machine learning techniques for flood forecasting. The techniques widely used for such forecasting include decision tree (DT) [21-23], ANN [15-17], MLC [24-26], fuzzy logic [18-20], and SOM [27-29]. These techniques contain high accuracy. Yet, big data has not been used much.

According to the literature review of big data use for flood forecasting or flood management in a research, it was basically used for data collection by rainfall and water level measurement devices, based on the internet of thing (IOT) [41]. The measurement values used included rainfall, water level, and water flow. Besides, meteorological data was also collected from the government website. Weather data and the one from NOAA Satellite was collected for flood forecasting. The forecasting technique used was K-mean by Holt-Winter's method, an unsupervised learning technique. For the results of error measurement of flood forecasting in the research, the errors of forecasted water level were measured on different dates in one, three, and six months by MAPE. The measurement in one month showed MAPE of 1.39% while the other months showed higher errors. Nonetheless, this error measurement of flood forecasting was to measure flood errors at a certain point only, not in different areas (only one tested area in the research). The highlight of the research was big data use for forecasting, resulting in high accuracy and low errors. For the analysis of the research limitations, big data used here was still data acquisition from traditional methods. No API development for automatic data downloading nor crowd source data use. Another research used big data for flood risk analytics and assessment [42-43]. The data used included crowd source data acquired from Twitter, rainfall detect sensors, and NASA. To download API for showing data on a website, users could see water level, rainfall, and moisture of soil in each particular area. Still, the research did not describe methods or techniques for flood risk assessment, but to notify users of flood monitoring through proposed data on the website. The highlight of the research was API development for automatic downloading. For the limitation, the presented results to users were not from forecasting each area whether or not floods happened. Users must use their own experiences to analyze or assess flood possibilities in the areas.

However, most research papers relying on machine learning for flood forecasting still use data collected by manpower or involved agencies in the traditional method, i.e., file recording or downloading by manpower, without automatic downloading. That is why they failed to conduct real-time flood forecasting. And it also affects the implementation cost as more labors must be hired, together with implementation delays. Fortunately, capabilities of software development technology today do not require manpower since API is developed and replaced manpower. API is a connecting channel between a website and another; between a user and a server; or from a server to a server. API is like a computer language that facilitates computers to communicate and exchange data freely to one another. It helps access to data or downloading data from data-serviced websites by API. Access scope to services to data can be automatically restricted with no need to use manpower to record data for future use. Data can be updated or downloaded in real-time. With such capabilities and qualities of API, involved service providers have increased in larger numbers. According to the studies

of API development for hydrology, developed API is used to assess construction/building damage caused by floods. To describe, users set dummy flood levels for damage assessment in order to see whether or not those exact levels will affect their constructions, and how many constructions will get affected. API development uses Python [44] and is offered for users to take data of elevation with high resolutions for flood modelling [45]. So far, API has not been used by involved agencies yet for downloading data of factors influencing floods in the form of big data for flood forecasting. According to the studies of related works, the advantages and disadvantages of flood forecasting can be concluded by different methods, as demonstrated in Table 1.

Table 1. Advantages and disadvantages of traditional methods and our proposed

Method	Advantages	Disadvantages
Machine Learning + Traditional Data	1. Flexible. Factors can be increased or reduced. High forecasting accuracy. 2. Possibly insufficient data for forecasting.	1. Budgets for more labors and time for data acquisition.
Hydrological Model + Traditional Data	Having been accepted and used for a long time. Reliable to be used for flood modeling.	Inflexible necessary factors. Without the data of a certain factor, the model will be unfunctional.
Our Proposed	1. Flexible. Factors can be increased or reduced. High forecasting accuracy.	1. Big data might require extra storage devices.
Machine Learning + Big Data	2. Data of factor can be downloaded automatically without manpower. Budgets are saved; with faster data for use.	2. Hi-speed central processing unit is required as it affects forecasting time.
	3. Sufficient data for forecasting.	

3 Methodology

3.1 Study areas

The study areas in this research were in the southern part of Thailand, including 2 provinces with regular floods, i.e., Surat Thani Province and Nakhon Si Thammarat Province. According to surveys of repeating flooded areas in the past 10 years by Department of Disaster Prevention and Mitigation (DDPM) and the Geo-Informatics and Space Technology Development Agency, Public Organization (GISTDA), there are the areas with regular repeating floods (8-10 times in 10 years), those with frequent floods (4-7 times in 10 years), and those with occasional floods (less than 3 times in 10 years). No matter what, flood forecasting methods in this research can definitely be applied in other areas.

This research used data from various sources with big data storage. The key quality of big data in this research was its high-volume. To elucidate, the data was big and in a large amount, which was available offline or online. This research used online data

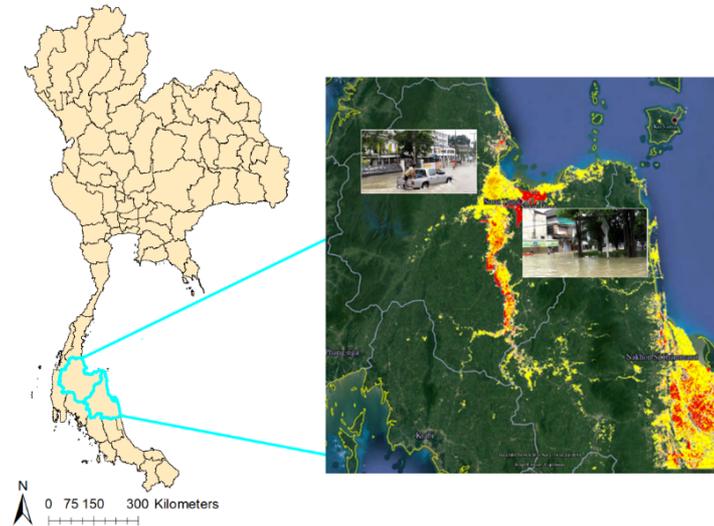


Fig. 1. Study Areas

through API downloading for real-time flood forecasting. Besides, data used was with high variety; in other words, it was various and either structured or unstructured. Furthermore, data used for flood forecasting was with high velocity. It changes all time and rapidly, with continuous transfer in real-time. This brought some limitations to manual data analysis. For example, rainfalls change with time. So, if the traditional methods if data transfer had been used, it would not have been possible to use rainfalls from measurement or forecasting for flood forecasting. On the other hand, it might have caused delays and could not have been used in time. Hence, this research proposed the development of data downloading in the form of API. Data used was divided into 3 main parts. Part 1 was meteorological and hydrological data, automatically downloaded from GLOFAS by API. This part included accumulated precipitation and probability forecasting or probability of precipitation at different levels, i.e., 50 mm, 150 mm, and 300 mm in each area, flood hazard 10-year return period, 5-year return period exceedance, and rainfall forecasting from big data of TMD. API was used to download data automatically at particular periods to be forecasted. Part 2 was geospatial data, downloaded from the databases of involved agencies through Web Feature Service (WFS). This part included height above sea level, slope, land use and land cover, 10-year repeating floods, and flow direction. Part 3 was crowdsourced data or volunteer data. This part was the written program for data acquisition from users in a certain area so that they can notify real situations in each area. This shall increase flood forecasting accuracy, because it is real-time data. Users can send data via mobile devices such as smart phones and tablets. For the details of the conceptual framework in this research, as demonstrated in Figure 2, each part was described about the title of the development of data downloading system in the form of API, flood forecasting using machine learning techniques, and accuracy assessment.

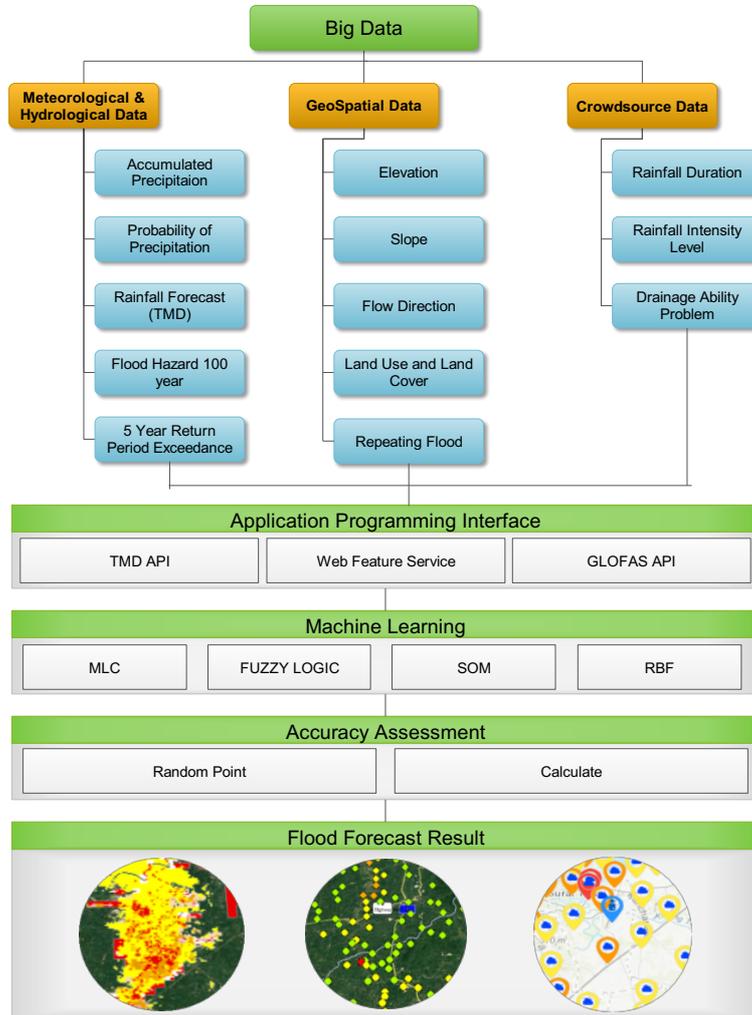


Fig. 2. Conceptual Framework

3.2 The development of application programming interface for flood forecasting

The development of data downloading system in the form of API in this research was divided into 3 main parts. Each part included different details. Part 1 was meteorological and hydrological data, automatically downloaded from GLOFAS and TMD. Downloading from GLOFAS included 5 factors, i.e., accumulated precipitation and probability forecasting or probability of precipitation at different levels, i.e., 50 mm, 150 mm, and 300 mm in each area, flood hazard 10-year return period, 5-year return period exceedance, and rainfall forecasting. PHP was used for the development. Code program was demonstrated in Figure 3. For downloading big data of TMD, it was to

download rainfall forecasting data that used PHP and Leaflet for the development. The example of code was demonstrated in Figure 4. Part 2 was downloaded geospatial data, including 5 factors, i.e., height above sea level, slope, land use and land cover, 10-year repeating floods, and flow direction. The development of this part relied on WFS because most data was stored in the form of Shapefile. It was different from Part 1, which was stored in the form of image. API was used as Web Map Service (WMS) instead. Part 3 was crowdsource data or volunteer data. This part included 3 factors, i.e., rainfall duration, rainfall intensity level, and drainage ability problem. PHP, Google Map, and API were used for the development in this part. The system downloaded data from these 3 developed parts. In addition, the data of all 13 factors were downloaded for flood forecasting. The details of each factor were demonstrated in Figure 5. The data of each factor was divided into multiple classes based on data qualities to brought for forecasting, to which machine learning techniques were used in the next step.

```
$curl = curl_init();
curl_setopt_array($curl, array(
    CURLOPT_URL =>
    "https://data.tmd.go.th/nwpapi/v1/forecast/location/daily/place?province=Surat Thani
    &amphoe=" . $samphur_data[$sam] . "&fields=tc,rh,rain,cond&date=" . date('Y-m-d') .
    "&duration=1&subarea=1",
    CURLOPT_RETURNTRANSFER => true,
    CURLOPT_ENCODING => "",
    CURLOPT_MAXREDIRS => 10,
    CURLOPT_TIMEOUT => 30,
    CURLOPT_HTTP_VERSION => CURL_HTTP_VERSION_1_1,
    CURLOPT_CUSTOMREQUEST => "GET",
    CURLOPT_HTTPHEADER => array(
        "accept: application/json",
        "authorization: #API_USER",
    ),
));
$response = curl_exec($curl);
$error = curl_error($curl);
curl_close($curl);
```

Fig. 3. TMD API

```
tilelayer[0] = L.tileLayer.wms("http://globalfloods-ows.ecmwf.int/glofas-ows/ows.py/?", {
    TIME: date,
    crs: L.CRS.EPSG4236,
    uppercase: true,
    layers: 'RPGM',
    format: 'image/png',
    transparent: true,
    version: '1.3.0',
    attribution: "Globalfloods.eu",
    opacity: opa,
    pane: 'tilePane',
    detectRetina: true,
    noWrap: true
}).addTo(map);
```

Fig. 4. GLOFAS API

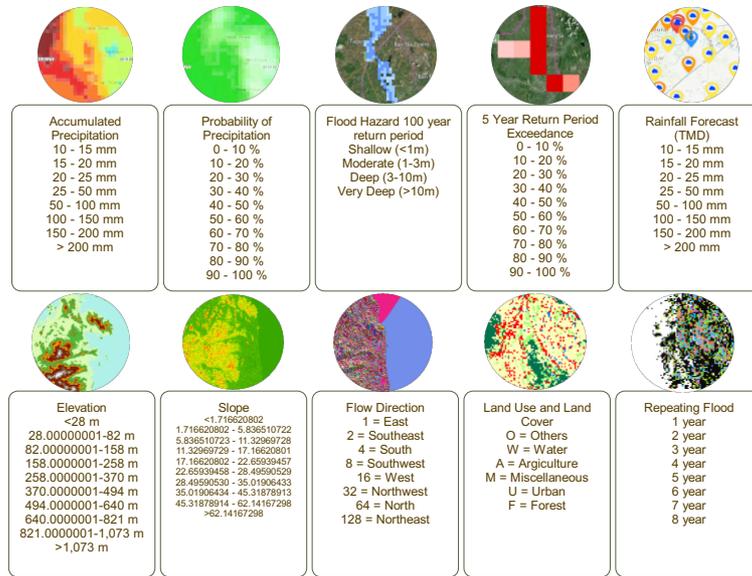


Fig. 5. Detail of Factors Data

3.3 Flood forecasting using machine learning techniques

At the step of flood forecasting by using each technique of machine learning for leading to learning process, the techniques used for comparison to find flood forecasting efficiency in this research included MLC, Fuzzy Logic, SOM, and RBF. The input factors were divided into 3 main parts, i.e., meteorological and hydrological Data, including 5 factors. Part 2 was geospatial data, including 5 factors. The last part was crowdsource data or volunteer data acquired from users for 3 factors. There were totally 13 input factors for learning and test as stated before under the title “The Development of Application Programming Interface for Flood Forecasting.” The results of forecasting in this research included 4 classes, i.e., non flood, low floods (below 20 cm), moderate floods (between 20-49 cm), and heavy floods (over 50 cm or equivalently). Then, the results were compared with the actual data from involved agencies. Field survey was also conducted for data collection. The study areas in this research included the areas in Surat Thani Province and Nakhon Si Thammarat Province.

Flood forecasting in this research contained the development of an algorithm for forecasting, as demonstrated in Table 2. The input data included 13 factors, of which results were from flooding forecasting in compliance with current coordinates of users. The first step was to read current coordinates of GPS from mobile devices. Next, TMD rainfalls were examined. If the coordinates were below 40 mm and the values of accumulated precipitation were below 4, the results of forecasting in that area showed “Non Flood.” But if now, the values of other factors must be downloading and brought for forecasting by machine learning in order to produce the result of forecasting at the location of users. The criteria of TMD Rainfalls and accumulated

precipitation were acquired from flood statistics analysis in the study areas in the past 10 years. It was found that means of rainfalls causing floods in each area were not below 40 mm/day. This research, therefore, applied the statistical values to the algorithm for flood forecasting so as to primarily screen non-flood areas. By doing so, forecasting would be faster.

Table 2. Flood forecast algorithm

Algorithm	Flood Forecast for User Location
Input:	UserLocation UL, RainfallTMD RT, AccPrecipitation AP, ProbPrecipitation PP, Flood100y FY, 5YearPE YE, Elevation E, Slope S, FlowDirection FD, LULC, RepeatFlood RF, RainfallDuration RD, RainfallIntensity RI, DrainageAbilityProblem DP
Output:	Flood Forecast Level of User Location (Non-Flood, Low Flood, Moderate Flood, Heavy Flood)
01:00	UL ← Get user location from mobile device //GeoLocation Command
02:00	RT ←TMD_API(UL) //Send user location for request rainfall forecast data)
03:00	IF RT < 40 THEN
04:00	AP (GLOFAS_API(UL)
05:00	IF AP < 4 THEN
06:00	FLUL (Non-Flood)
07:00	ELSE
08:00	GO TO 10
09:00	ELSE
10:00	AP ← GLOFAS_API(UL)
11:00	PP ← GLOFAS_API(UL)
12:00	FY ← GLOFAS_API(UL)
13:00	YE ← GLOFAS_API(UL)
14:00	E ← GEODB_API(UL)
15:00	S ← GEODB_API(UL)
16:00	FD ← GEODB_API(UL)
17:00	LL ← GEODB_API(UL)
18:00	RF ← GEODB_API(UL)
19:00	RD ← GLOFAS_API(UL)
20:00	RI ← GLOFAS_API(UL)
21:00	DP ← GLOFAS_API(UL)
22:00	FLUL ← MLC_M (RT, AP, PP, FY, YE, E, S, FD, LL, RF, RD, RI, DP)
23:00	RETURN FLUL

3.4 Accuracy assessment

This research relied on accuracy assessment, of which correctly classified instances were calculated. Data used to compare with the results of flood forecasting by MLC, fuzzy logic, SOM, and RBF was the data from the past floods. That data was acquired from field survey by the researchers and DDPM. The equation for accuracy calculation was demonstrated in Equation (1). Data used for accuracy assessment included 1,000 points in the study areas. The number of those points were acquired from automatic point randomization process. Apart from this, this research also conducted forecasting error assessment by using mean absolute percent error (MAPE). The equation for calculation was demonstrated in Equation (2).

$$\text{Correctly Classified Instances} = (\text{True forecasting})/(\text{Total Forecasting}) \quad (1)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|\hat{y}_t - y_t|}{y_t} \times 100 \quad (2)$$

where n is the size of the sample, y_t is the value forecasted by the model for time point t , and y_t is the value observed at time point t .

4 Result and Discussion

The results of this research were divided into 2 parts, i.e., the results of API development for flood forecasting and of flood forecasting. The first part referred to the results of system development to download data from GLOFAS and TMD Big Data, as demonstrated in Figure 6 – 8. According to Figure 6, data of accumulated precipitation between 3 – 6 January 2019 was downloaded. This implied different accumulated precipitation each day, represented by the colors of pixel for acknowledgement of accumulated precipitation in each particular area. Periods could be chosen for data use, starting from 1 January 2011 to the present date. So, besides using accumulated precipitation for flood forecasting, it can also be used for analysis in other aspects. According to Figure 7, the results of downloading probability forecasting or probability of precipitation at 150 mm was demonstrated. The data was downloaded from GLOFAS in the form of Raster, supporting analysis capability by using the programs called GIS, remote sensing, and image processing. According to Figure 8, the data of forecasted rainfalls from TMD was downloaded. Rainfalls were represented by different colors of pins containing different meanings. To explain, yellow pins referred to rainfalls below 40 mm. Orange pins referred to rainfalls between 41 – 80 mm. And red pins referred to rainfalls over 81 mm. The demonstration of rainfalls could be chosen based on exact dates to be used.

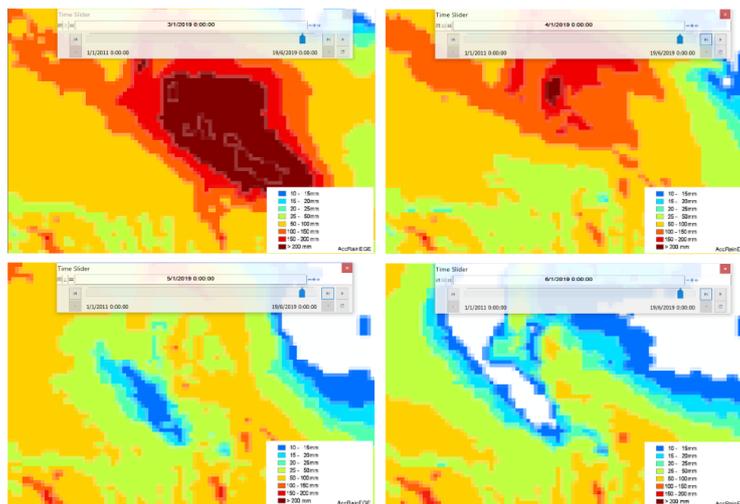


Fig. 6. Example of Accumulated Precipitation Data from GLOFAS API

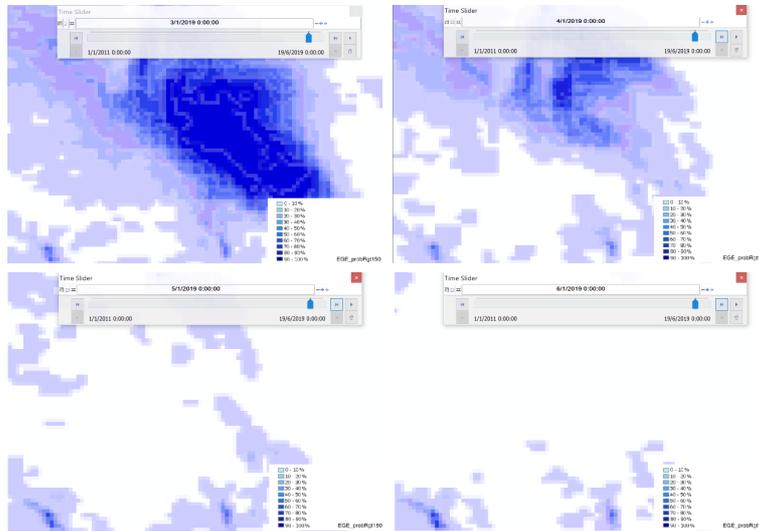


Fig. 7. Example of Probability of Precipitation Data from GLOFAS API

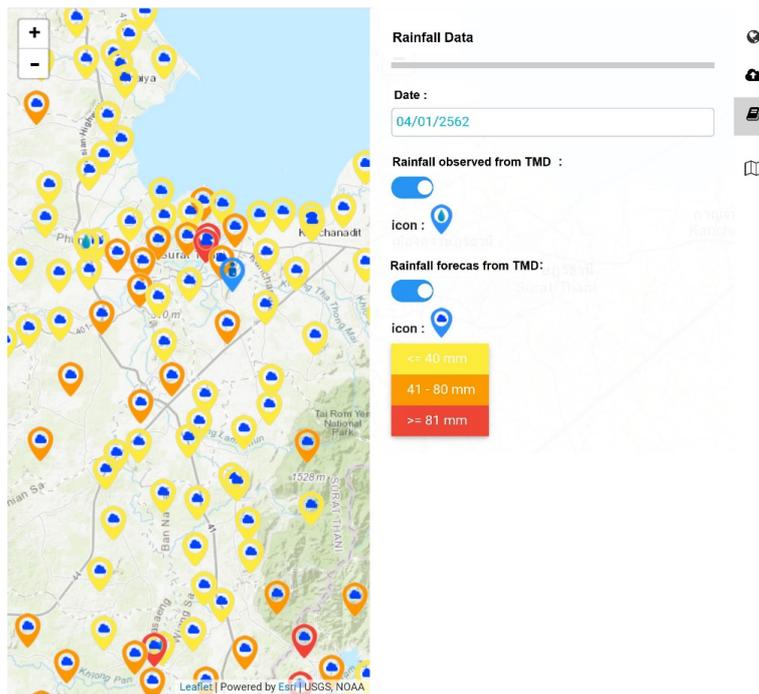


Fig. 8. Example of Rainfall Data from TMD

The results of flood forecasting in this research included 4 machine learning techniques, i.e., MLC, fuzzy logic, SOM, and RBF in the areas of Surat Thani Province

and Nakhon Si Thammarat Province. Such results in each areas were demonstrated in Figure 9-10. According to Figure 9, the results in Surat Thani Province from each method were demonstrated in Figure 9(a) - 9(d), by MLC, fuzzy logic, SOM, and RBF, respectively; and demonstrated in pixel. Forecasting was represented by different colors, including 4 classes. To illustrate, non flood was represented by green; low floods (water levels below 20 cm) was represented by yellow; moderate floods (water levels between 21 – 49 cm) was represented by orange; and heavy floods (water level above 50 cm or equivalently) was represented by red. According to forecasting analysis by visual inspection, MLC, fuzzy logic, and SOM produced similar results of forecasting. To elaborate, most areas in Surat Thani Province showed “non flood” and most flooded areas showed low floods. Areas with moderate and heavy floods had similar sizes. Forecasting by those 3 techniques produced different results from those by RBF, which showed more flooded areas than other techniques. Such additional areas belonged to Khiri Rat Nikhom District, with heavy floods. According to the results of forecasting in the areas of Nakhon Si Thammarat Province in Figure 10(a) - 10(d), the results conformed with those of Surat Thani Province, that is, MLC, fuzzy logic, and SOM produced similar results of forecasting. To elaborate, most areas in Nakhon Si Thammarat Province showed “non flood.” MLC, fuzzy logic, and SOM produced similar results of forecasting and the same as Surat Thani Province; whereas RBF produced the different results from those by other techniques. According to the overall results of flood forecasting in both provinces, assessed by visual inspection, the results produced by MLC, fuzzy logic, and SOM were similar; whereas RBF produced the different results. Nevertheless, for accuracy assessment of flood forecasting by different techniques in order to compare most accurate techniques, 1,000 points in the study areas were randomized and compared with flood situations in the real areas. The results were demonstrated in Figure 11 – 13. Those in Figure 11(a) - (d) were from flood forecasting in Surat Thani Province by MLC, fuzzy logic, SOM, and RBF, respectively. Those in Figure 12(a) - (d) were from flood forecasting in Nakhon Si Thammarat Province by MLC, fuzzy logic, SOM, and RBF, respectively. The results demonstrated in both provinces were represented by points. Each point was the coordinate of a particular location or area. The results of forecasting were represented by the color of each point. There were 4 classes all together, i.e., non flood was represented by green; low floods (water levels below 20 cm) was represented by yellow; moderate flood (water levels between 21 – 49 cm) was represented by orange; and heavy floods (water level above 50 cm or equivalently) was represented by red. According to accuracy assessment of flood forecasting by MLC, fuzzy logic, SOM, and RBF, the results were demonstrated in Figure 13(a) - (d) in the form of point. The accurate ones (true forecasting) were represented by green. The inaccurate ones (false forecasting) were represented by red. According to the analysis of the figures, the method with the least false forecasting was MLC. The areas with the falsest forecasting were Muang District and Kanchanadit District, Surat Thani Province. The second least false forecasting were fuzzy logic, SOM, and RBF. By fuzzy logic, The area with the falsest forecasting was Pak Phanang District, Nakhon Si Thammarat Province; the same results as those by SOM. RBF produced huge false forecasting in many areas, i.e., Phun Phin District, Khiri Rat Nikhom District, and Khian Sa District in

Surat Thani Province; and Muang District, Ron Phibun District, Hua Sai District, Charoem Phakiat Distric, and Chian Yai District in Nakhon Si Thammarat Province.

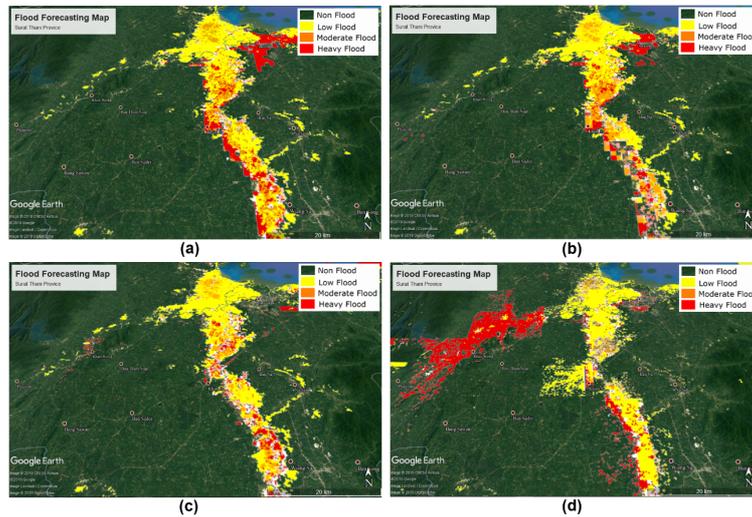


Fig. 9. Flood forecasting result in Surat Thani province using techniques a) MLC b) Fuzzy Logic c) SOM d) RBF.

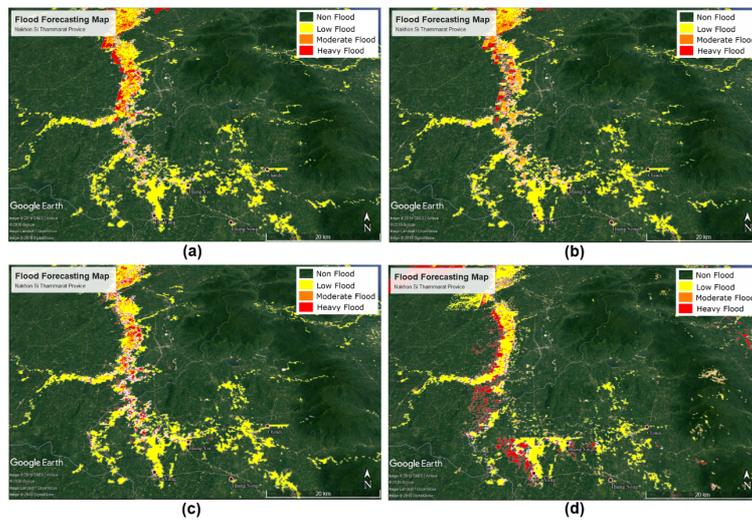


Fig. 10. Flood forecasting result in Nakhon Si Thammarat province using techniques a) MLC b) Fuzzy Logic c) SOM d) RBF

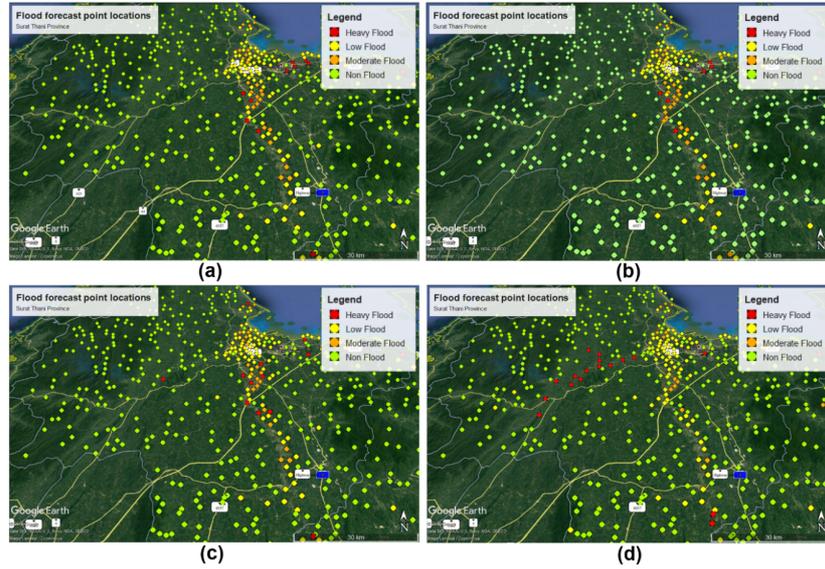


Fig. 11. Flood forecasting result in Surat Thani province using techniques a) MLC b) Fuzzy Logic c) SOM d) RBF

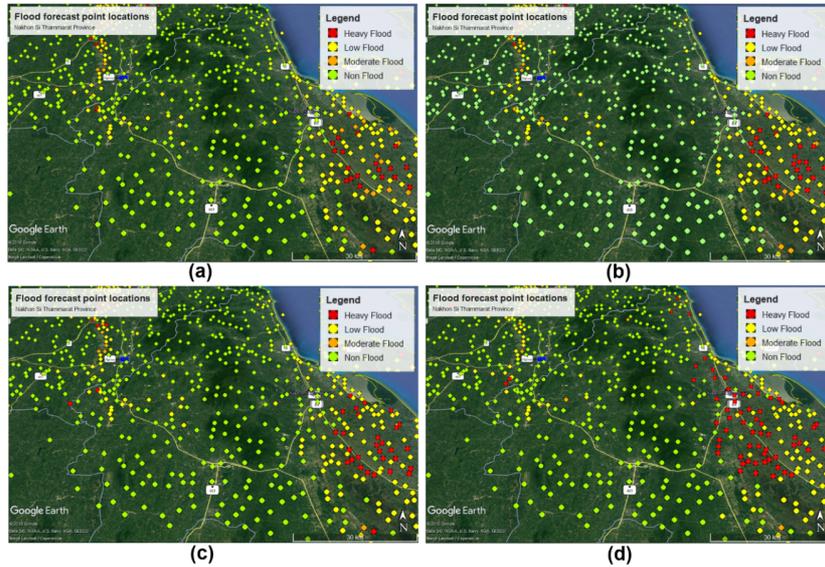


Fig. 12. Flood forecasting result in Nakhon Si Thammarat province using techniques a) MLC b) Fuzzy Logic c) SOM d) RBF

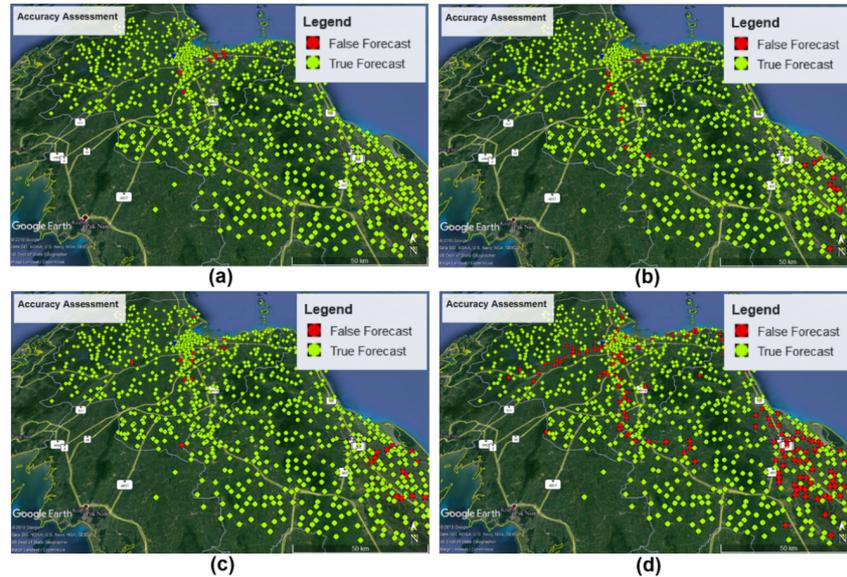


Fig. 13. Accuracy Assessment using a) MLC b) Fuzzy c) SOM and d) RBF

Regarding accuracy assessment of flood forecasting by the 4 techniques, as demonstrated in Figure 14(a). The results of the assessment revealed that they were in line with visual inspection. The most accurate technique was MLC (99.2%), followed by fuzzy logic (97.8%), SOM (96.6%), and RBF (83.3%), respectively. What’s more, this research also conduct forecasting error assessment by using MAPE. The results of assessment were demonstrated in Figure 14(b). The technique with the least error was MLC (2.13%), followed by fuzzy logic (2.31%), SOM (2.88%), and RBF (19.85%), respectively. When analyzing errors from each technique based on the standard interpretation criteria, it was fixed that MAPE below 10 referred to highly accurate forecasting. Between 10 – 20 referred to good forecasting. Between 20-50 referred to reasonable forecasting. And over 50 referred to inaccurate forecasting. Thus, when considering the results of forecasting in this research, the 3 techniques, i.e., MLC, fuzzy logic, and SOM showed highly accurate forecasting or very low errors. RBF was the only 1 technique that showed “good forecasting.”

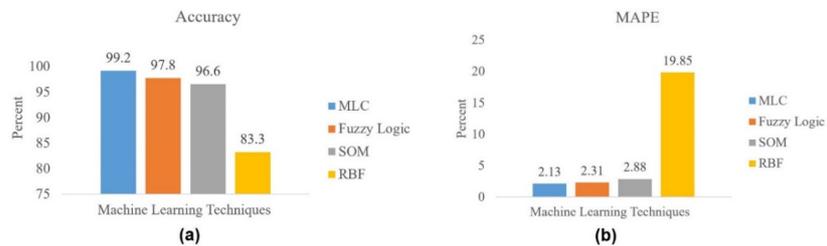


Fig. 14. Flood forecasting results validation a) Accuracy b) MAPE

This research was implemented with the survey of real flood data in each particular area. Field survey was conducted and data was brought from involved agencies. Crowdsourcing data was also taken into account from their flood notifications on thaiflood.org, which the researchers had developed for users to notify flood situations in each area in order to use data for flood management in the area and to examine flood forecasting in the research. The examples of real flooded areas were demonstrated in Figure 15. The points with the red flags represented real heavy floods. This was in accordance with the results of flood forecasting by the methods presented in this research. Likewise, the points with the orange flags represented real moderate floods; and the points with the yellow flags represented real low floods.



Fig. 15. Flood forecasting result overlaid with ground truth data

5 Conclusion

This research developed API for downloading big data automatically. It included factors influencing floods and was downloaded from involved agencies in order to be used for flood forecasting in the areas of Surat Thani Province and Nakhon Si Thammarat Province, Thailand. 4 machine learning techniques were used, i.e., MLC, fuzzy logic, SOM, and RBF. The results of flood forecasting revealed that using big data with machine learning produced highly accurate forecasting. The most accurate technique was MLC. Additionally, the methods presented in this research helped flood forecasting process become faster, with better advantages than traditional method with traditional data. They also helped saving cost in term of labors and time to collect or download data for forecasting.

The future research to be implemented is the development of flood forecasting systems for real-time forecasting and demonstrating results in each area in the form of a web application or a mobile application. The purposes are to allow users to keep follow-up and get ready to cope with upcoming floods; and to provide supporting device-

es for flood warning to people and involved person. As a consequence, flood management will be done more efficiently and damage will be reduced.

6 References

- [1] A. A. Monrat, R. U. Islam, M. S. Hossain, K. Andersson, “Challenges and Opportunities of Using Big Data for Assessing Flood Risks.,” In Applications of Big Data Analytics, pp. 31-42, 2018. https://doi.org/10.1007/978-3-319-76472-6_2
- [2] M. Brilly, M. Polic, “Public perception of flood risks, flood forecasting and mitigation.,” Natural Hazards and Earth System Science, vol.5, no. 3, pp. 345-355. 2005. <https://doi.org/10.5194/nhess-5-345-2005>
- [3] A. Oussous, F. Z. Benjelloun, A. A. Lahcen, S. Belfkih, “Big Data technologies: A survey.,” Journal of King Saud University-Computer and Information Sciences, Vol. 30, no. 4, pp. 431-448, 2018. <https://doi.org/10.1016/j.jksuci.2017.06.001>
- [4] A. K. Bhadani, D. Jothimani, “Big data: challenges, opportunities, and realities.,” In Effective Big Data management and opportunities for implementation, pp. 1-24, 2016. <https://doi.org/10.4018/978-1-5225-0182-4.ch001>
- [5] G. Kan, X. He, L. Ding, J. Li, K. Liang, Y. Hong, “Study on applicability of conceptual hydrological models for flood forecasting in humid, semi-humid semi-arid and arid basins in China.,” Water, Vol. 9, no. 10, pp. 719, 2017. <https://doi.org/10.3390/w9100719>
- [6] G. Tayfur, V. Singh, T. Moramarco, S. Barbetta, “Flood hydrograph prediction using machine learning methods.,” Water, Vol. 10, no. 8, pp. 968, 2018. <https://doi.org/10.3390/w10080968>
- [7] H. E. N. R. I. K. Madsen, D. Rosbjerg, J. Damgard, F. S. Hansen, “Data assimilation in the MIKE 11 Flood Forecasting system using Kalman filtering.,” International Association of Hydrological Sciences, Publication, Vol. 281, pp. 75-81, 2003
- [8] M. M. Rahman, N. K. Goel, D. S. Arya, “Development of the Jamuneswari flood forecasting system: Case study in Bangladesh.,” Journal of Hydrologic Engineering, vol. 17, no. 10, pp. 1123-1140, 2011. [https://doi.org/10.1061/\(asce\)he.1943-5584.0000565](https://doi.org/10.1061/(asce)he.1943-5584.0000565)
- [9] P. Kadam, D. Sen, “Flood inundation simulation in Ajoy River using MIKE-FLOOD.,” ISH Journal of Hydraulic Engineering, Vol. 18, no. 2, pp. 129-141, 2012. <https://doi.org/10.1080/09715010.2012.695449>
- [10] M. Islam, N. Hofstra, E. Sokolova, “Modelling the Present and Future Water Level and Discharge of the Tidal Betna River.,” Geosciences, Vol. 8, no. 8, pp. 271, 2018. <https://doi.org/10.3390/geosciences8080271>
- [11] F. E. Hicks, T. Peacock, “Suitability of HEC-RAS for flood forecasting.,” Canadian Water Resources Journal, Vol. 30, no. 2, pp. 159-174, 2005. <https://doi.org/10.4296/cwrj3002159>
- [12] D. Mai, F. De Smedt, “A combined hydrological and hydraulic model for flood prediction in Vietnam applied to the huong river basin as a test case study.,” Water, Vol. 9, no. 11, pp. 879, 2017. <https://doi.org/10.3390/w9110879>
- [13] K. Solaimani, “Flood forecasting based on geographical information system.,” African Journal of Agricultural Research, Vol. 4, no. 10, pp. 950-956, 2009
- [14] B. Thakur, R. Parajuli, A. Kalra, S. Ahmad, R. Gupta, “Coupling HEC-RAS and HEC-HMS in Precipitation Runoff Modelling and Evaluating Flood Plain Inundation Map.,” World Environmental and Water Resources Congress, pp. 240-251. 2017. <https://doi.org/10.1061/9780784480625.022>

- [15] S. H. Elsafi, "Artificial neural networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan.," Alexandria Engineering Journal, Vol. 53, no. 3, pp. 655-662, 2014. <https://doi.org/10.1016/j.aej.2014.06.010>
- [16] Y. Tahmasebi Biragani, F. Yazdandoost, H. Ghalkhani, "Flood Forecasting Using Artificial Neural Networks: an Application of Multi-Model Data Fusion Technique.," Journal of Hydraulic Structures, Vol. 2, no. 2, pp. 62-73, 2016
- [17] A. Jabbari, D. H. Bae, "Application of Artificial Neural Networks for Accuracy Enhancements of Real-Time Flood Forecasting in the Imjin Basin.," Water, Vol. 10, no. 11, pp. 1626, 2018. <https://doi.org/10.3390/w10111626>
- [18] E. D. P. Perera, L. Lahat, "Fuzzy logic based flood forecasting model for the Kelantan River basin, Malaysia.," Journal of hydro-environment research, Vol. 9, no. 4, pp. 542-553, 2015. <https://doi.org/10.1016/j.jher.2014.12.001>
- [19] A. K. Lohani, N. K. Goel, K. K. S. Bhatia, "Improving real time flood forecasting using fuzzy inference system." Journal of Hydrology, Vol. 509, pp. 25-41, 2014. <https://doi.org/10.1016/j.jhydrol.2013.11.021>
- [20] L. C. Chang, F. J. Chang, Y. H. Tsai, "Fuzzy exemplar based inference system for flood forecasting.," Water Resources Research, Vol. 41, no. 2, pp. 1-12, 2005. <https://doi.org/10.1029/2004wr003037>
- [21] J. Noymanee, N. O. Nikitin, A. V. Kalyuzhnaya, "Urban Pluvial Flood Forecasting using Open Data with Machine Learning Techniques in Pattani Basin.," Procedia computer science, Vol. 119, pp. 288-297, 2017. <https://doi.org/10.1016/j.procs.2017.11.187>
- [22] M. S. Tehrani, B. Pradhan, M. N. Jebur, "Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS.," Journal of Hydrology, Vol. 504, pp. 69-79, 2013. <https://doi.org/10.1016/j.jhydrol.2013.09.034>
- [23] B. Ramaswamy, L. P. PB, K. Vishruth, "Urban Flood Forecast using Machine Learning on Real Time Sensor Data.," Transactions on Machine Learning and Artificial Intelligence, Vol. 5, no. 5, pp. 69, 2017
- [24] Y. U. Badaru, O. S. M. Y. I. Ishiaku, Y. M. Nassir, "Analysis of Sensor Imaging and Field-Validation for Monitoring, Evaluation and Control Future Flood Prone Areas along River Niger and Benue Confluence Ecology, Lokoja, Nigeria.," environment, Vol. 4, no. 22, pp. 90-99, 2014
- [25] A. Heidari, B. Saghafian, R. Maknoon, "Assessment of flood forecasting lead time based on generalized likelihood uncertainty estimation approach.," Stochastic Environmental Research and Risk Assessment, Vol. 20, no. 5, pp. 363-380, 2006. <https://doi.org/10.1007/s00477-006-0032-y>
- [26] L. Alfieri, M. Berenguer, V. Knechtel, K. Liechti, D. Sempere-Torres, M. Zappa, "Flash flood forecasting based on rainfall thresholds.," Handbook of Hydrometeorological Ensemble Forecasting, pp. 1223-1260, 2019. https://doi.org/10.1007/978-3-642-39925-1_49
- [27] L. C. Chang, M. Amin, S. N. Yang, F. J. Chang, "Building ANN-based regional multi-step-ahead flood inundation forecast models.," Water, Vol. 10, no. 9, pp. 1283, 2018. <https://doi.org/10.3390/w10091283>
- [28] I. Valença, T. Ludermit, "Hybrid Systems for River Flood Forecasting Using MLP, SOM and Fuzzy Systems.," In International Conference on Artificial Neural Networks, pp. 557-566, 2009. https://doi.org/10.1007/978-3-642-04274-4_58
- [29] M. Herbst, M. C. Casper, J. Grundmann, O. Buchholz, "Comparative analysis of model behaviour for flood prediction purposes using Self-Organizing Maps.," Natural Hazards and Earth System Sciences, Vol. 9, no. 2, pp. 373-392, 2009. <https://doi.org/10.5194/nhess-9-373-2009>

- [30] A. Anele, Y. Hamam, A. Abu-Mahfouz, E. Todini, "Overview, comparative assessment and recommendations of forecasting models for short-term water demand prediction.," *Water*, Vol. 9, no. 11, pp. 887, 2017. <https://doi.org/10.3390/w9110887>
- [31] A. Mosavi, P. Ozturk, K. W. Chau, "Flood Prediction Using Machine Learning Models: Literature Review.," *Water*, Vol. 10, no. 11, pp. 1536, 2018. <https://doi.org/10.3390/w10111536>
- [32] A. Jabbari, J.-M. So, D.H. Bae, "Accuracy assessment of real-time flood forecasting of coupled hydrological and mesoscale meteorological models.," *Nat. Hazards Earth Syst. Sci. Discuss*, <https://doi.org/10.5194/nhess-2017-447>, 2018. <https://doi.org/10.5194/nhess-2017-447>
- [33] S. Jeon, B. Hong, J. Kim, H. J. Lee, "Stock Price Prediction based on Stock Big Data and Pattern Graph Analysis.," In *IoTBD*, pp. 223-231, 2016. <https://doi.org/10.5220/0005876102230231>
- [34] G. V. Attigeri, M. P. MM, R. M. Pai, A. Nayak, "Stock market prediction: A big data approach.," In *TENCON 2015-2015 IEEE Region 10 Conference*, pp. 1-5, 2015. <https://doi.org/10.1109/tencon.2015.7373006>
- [35] G. Thangarasu, K. Subramanian, "Big Data Analytics for Improved Care Delivery in the Healthcare Industry.," *International Journal of Online Engineering*, Vol. 15, no. 10, pp. 4-19, 2019. <https://doi.org/10.3991/ijoe.v15i10.10875>
- [36] K. Adam, M. A. Majid, M. A. I. Fakhherldin, J. M. Zain, "A Big Data Prediction Framework for Weather Forecast Using MapReduce Algorithm.," *Advanced Science Letters*, Vol. 23, no. 11, pp. 11138-11143, 2017. <https://doi.org/10.1166/asl.2017.10237>
- [37] M. Yu, C. Yang, Y. Li, "Big data in natural disaster management: a review.," *Geosciences*, Vol. 8, no. 5, pp. 165, 2018
- [38] J. Wang, X. Yang, "An Automatic Online Disaster Monitoring Network: Network Architecture and a Case Study Monitoring Slope Stability.," *International Journal of Online Engineering*, Vol. 14, no. 3, pp. 40-51, 2018. <https://doi.org/10.3991/ijoe.v14i03.7669>
- [39] S. Dhamodaran, K. R. Sachin, R. Kumar, "Big data implementation of natural disaster monitoring and alerting system in real time social network using hadoop technology.," *Indian Journal of Science and Technology*, Vol. 8, no. 22, pp. 1, 2015. <https://doi.org/10.17485/ijst/2015/v8i22/79102>
- [40] L. Alfieri, P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen, F. Pappenberger, "GloFAS-global ensemble streamflow forecasting and flood early warning.," *Hydrology and Earth System Sciences*, Vol. 17, no. 3, pp. 1161, 2013. <https://doi.org/10.5194/hess-17-1161-2013>
- [41] S. K. Sood, R. Sandhu, K. Singla, V. Chang, "IoT, big data and HPC based smart flood management framework.," *Sustainable Computing: Informatics and Systems*, Vol. 20, pp. 102-117, 2018. <https://doi.org/10.1016/j.suscom.2017.12.001>
- [42] A. A. Monrat, R. U. Islam, M. S. Hossain, K. Andersson, "Challenges and Opportunities of Using Big Data for Assessing Flood Risks.," In *Applications of Big Data Analytics*, pp. 31-42, 2019. https://doi.org/10.1007/978-3-319-76472-6_2
- [43] T. D. Indriasari, K. Anindito, E. B. Julianto, "A Pangaribuan, Mobile and Web Application for Mapping Disaster Volunteers' Position in Indonesia.," *International Journal of Interactive Mobile Technologies*, Vol. 11, no. 3, pp. 98-112, 2017. <https://doi.org/10.3991/ijim.v11i3.6477>
- [44] H. McGrath, M. Kotsollaris, E. Stefanakis, M. Nastev. "Flood damage calculations via a RESTful API.," *International Journal of Disaster Risk Reduction*, Vol. 35, pp. 1-8, 2019. <https://doi.org/10.1016/j.ijdrr.2019.101071>

- [45] H. McGrath, E. Stefanakis, M. Nastev, “DEM Fusion of Elevation REST API Data in Support of Rapid Flood Modelling.,” *GEOMATICA*, Vol. 70, no. 4, pp. 283–297, 2016. <https://doi.org/10.5623/cig2016-402>

7 Acknowledgment

This research project was financially supported by the Agricultural Research Development Agency Public Organization, National Research Council of Thailand, and Thailand Research Organizations Network. The authors would like to thank the Geoinformatics and Space Technology Development Agency, Public Organization, Thailand Meteorological Department, Department of Disaster Prevention and Mitigation, Google Inc. and the GLOFAS website for providing the remotely sensed data (Meteorological and Hydrological Data) used in preparation of the paper.

8 Authors

Supattra Puttinaovarat is a PhD lecturer at Faculty of Science and Industrial Technology, Prince of Songkla University, Surat Thani Campus. Her research interest includes Geographic Information System and Information Technology.

Paramate Horkaew is an assistant professor at the School of Computer Engineering, Suranaree University of Technology, Thailand. His main research interests include Computational Anatomy, Digital Geometry Processing, Computer Vision and Graphics. phorkaew@sut.ac.th

Article submitted 2019-07-09. Resubmitted 2019-08-25. Final acceptance 2019-08-26. Final version published as submitted by the authors.