Journal of ENVIRONMENTAL Treatment Techniques

J. Environ. Treat. Tech.
ISSN: 2309-1185

Journal web link: http://www.jett.dormaj.com

Data Understanding for Flash Flood Prediction in Urban Areas

Nur Shuhada Abdul Malek [1], Syamil Zayid [2], Zaifulasraf Ahmad [3], Suraya Ya'acob [4*], Nur Azaliah Abu Bakar [5]

Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia

Received: 19/11/2019 Accepted: 18/02/2020 Published: 20/05/2020

Abstract

Flash flood has become one of the major disastrous events, especially in urban areas in Malaysia. It has become more prominent to city dwellers, causing massive loss of infrastructures, damage to people, and disruption in business and daily activities. Population growth and rapid development of urban areas have worsened the situation even more. Since the era of Big Data, the possibility to analyse complex data coming from heterogeneous sources, which can be used to predict flash flood, has given a different perspective and hope for finding innovative ways to reduce the impact of flood, especially in urban areas. The purpose of this study is to understand data needed to produce predictive visual analytics for flash flood forecasting using Cross-Industry Standard Process for Data Mining (CRISP-DM) Methodology. Focusing on understanding the flash flood data, this paper intends to characterize data pertaining to disaster management and identify the right data that can facilitate more accurate decision making by stakeholders. Literature review was done to determine which data are needed in the Malaysian urban setting. The research found the critical factors for determining flash flood occurrence in Malaysia are unique due to the tropical climate and urbanization. Therefore, it is important to understand and characterize these factors for more effective and accurate data collection and predictive analytics later. Based on the findings, the most significant factors identified for flash flood prediction are rainfall, urbanization, and fluvial flood which eventually lead to blocked drainage. Details of data under these categories will be analysed as part of data understanding of flash flood occurrence. This study intends to uncover the potential of using Predictive Visual Analytics in flood forecasting and also to discuss how prediction can bring values to the Malaysian environment and create a sustainable ecosystem.

Keywords: Flash Flood, Disaster, Predictive Analytics, Data Understanding

1 Introduction

Flash flood is a sudden and rapid flooding of low-lying areas. It is one of the hazards that may cause loss of life, injury, and destroyed or damaged assets, which could affect the society, economy, and environment (1). Due to monsoon season and heavy rainfall, Malaysia is exposed to the flash flood risk every year. This condition becomes worse when rapid urbanization takes place. As an example, Kuala Lumpur is well known as the capital and the largest city in Malaysia. According to Department of statistic Malaysia, the population of Kuala Lumpur is estimated at 1.78 million as of 2019. It is among the fastest growing metropolitan regions in South-East Asia in terms of population, economic, and social development. Nevertheless, in terms of climate, Kuala Lumpur is one of the equatorial regions often beset by prolonged rainfall and storms that eventually lead to flood.

Several factors cause flash flood situation. Specifically, in case of Kuala Lumpur that is a big city located in a tropical country, the research found four factors contributing to flash flood occurrence: i) rainfall and climatic changes, ii) urban changes and anthropogenic activities, iii) network and catchment factors, and iv) geomorphological features. According to (1), the climatic changes have a major impact on the rainfall intensity. The climate is one of the major factors that will lead to the rainfall intensity which has the potential to cause flood. In relation to the flash flood issue in Malaysia, this research intends to implement the big data analytic approach in prediction of flood using the occurrence pattern of flash flood in Malaysia. Big data has a good potential to play an important role in generating knowledgeable information that ignite business and government interest in driving towards better decision making and mitigation plan for flood situation. This is the case when paying attention to findings of (2) indicating that the inefficient solid waste management is also one of the contributors to flood situation.

Big Data usually known as 5Vs which are volume, variety, velocity, veracity, and value. As for big data

Corresponding author: Suraya Ya'acob, Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia. E-mail: suraya.yaacob@utm.my.

definition, these 5Vs can be translated as a big volume of data collection that has velocity of complexities from variety instruments and sources which has the potential veracity and value to be used for data-driven decision makers in the organizations. To discover the required data and to get information, data analysis approaches such as descriptive, predictive, and prescriptive are used. Therefore, with the result obtained from the analysis, data are used to generate useful information supporting better decision making. As for this research, the implementation of Big Data Analytics is to predict the flash flood occurrence supports sustainability of cities and community in Malaysia. Therefore, the present study intends to understand and characterize data pertaining to flood situations and identify the right data that can facilitate more accurate decision making regarding to flash flood situation. At the end, it can help to reduce flash flood impact on society and economy in urban areas such as Kuala Lumpur.

1.1 More demand for Flash Flood Prediction.

In (3), it was found that, in recent years, more researchers have started to explore and integrate the predictive analytics techniques of data exploration into visual analytics systems. The capabilities of predictive analytics in reducing risk, making more intelligent decisions, and generating different customer experiences have attracted a lot of industrial players implement it in their business (4). Another study (5) investigated how the visual analytics community has recently focused on building interactive visualizations and associating them with predictive analytics methods. It was achieved through translating real data into the knowledge for their business decisions. The descriptive and predictive analytics are among four categories from business analytics. They differ in the way the use data: the descriptive analytics summarizes what has happened by collecting relevant data, storing them, and presenting the information to find the trends (4); on the other hand, the predictive analytics goes beyond that; it purposely provides insight into what and why will happen in the future by analysing the current and historical data (4).

Predictive analytics uses machine learning algorithms and statistical analysis techniques to predict flash flood future trends (36). In more details, the authors in (6) found that predictive analytics can provide the organization or business with a better understanding of what people need and, at the same time, helps to identify potential mitigation plans.

In (37), an interactive visual analytics application was proposed to be combined with automatic predictive visual analytics supported by domain experts to investigate the environmental conditions. The automatic predictive analysis has achieved results of a high level of accuracy (5). A predictive model can forecast the buying patterns, potential risks, and possible prospects through providing a deep understanding of the customers (38), and according to (39), it is capable of foreseeing possible future stories from the past similar cases via a predictive analysis system. The researchers in (40) mentioned that, in the current era of big data, machine learning, and Artificial Intelligence (AI), many visual analytics systems have been recently used to produce different predictive models applicable to defining unseen data or information.

1.2 Impact of Flash Flood Prediction

The application of big data to the prediction of flood can reduce many hazards to the environment, society, and economy. An early identification of hazards will help to manage residual risks. Authorities can formulate more comprehensive mitigation plans by taking into consideration three different entities: i) citizens affected by the flash flood, ii) the agencies that are expected to manage the flash flood, and iii) the risk reduction experts. At the end, the initiative practically taken can reduce risk of flash flood to the economy, people, process, technology, society, ecology, and environment in Malaysia.

In brief, the research objective is to reduce flash flood risk in Kuala Lumpur. In order to do that, the researchers intend to develop predictive visual analytics based on the environment challenges as shown in Table 1.

Table 1: Research Summary

Elements	Description
What	Developing predictive visual analytics to reduce flash flood risk in Kuala Lumpur, Malaysia.
Why	Flash flood has risk of loss of life, injury, and destroyed or damaged assets, which could occur to society, economy, and environment.
When	Rainy season throughout the year.
Where	Malaysia and, more specifically, the urban areas. In this case the research focuses on Kuala Lumpur as a capital city and Selangor as most developed state in Malaysia.
Who	Affected environment and economy, urban regions, people's health, and business and social activities.
How	To identify and characterize factors that contribute to the flash flood in the Malaysian environment.

Due to the Cross-Industry Standard Process for Data Mining (CRISP) methodology, there are six phases to implement predictive analytics. Due to environment-related content, this paper will focus more on phase 2, i.e., data understanding. The research found the importance of data understanding as to get the knowledge about the data, the needs that the data will satisfy, the availability, the requirements and the sources of the data. Furthermore, data understanding is important to connect between the flash-flood environment context and analytical preparation need to be done later. In data understanding phase, this study will identify potential of environmental data to be used; then, it characterizes and describes each data for predictive modelling in the next phase. The rest of the present paper is organized as follow. Section 1 introduces the research as a whole. Section 2 provide a background of the study

explaining the flash flood environment and discussing the predictive analytics. Section 3 explains the methodology used for flash flood data understanding. Section 4 discusses the factors identified as the gist of this paper. Finally, section 5 provides the conclusion and the summary of future work.

2 Working Background

2.1 Flash Flood Situation in Malaysia

Flash flood is a global issue faced by major cities around the world. It occurs because of a complex combination of meteorological and hydrological extremes such as extreme precipitation and flows. Flash flood can also be the result of human activities, including unplanned growth and development in floodplains, or it can take place due to the breach of a dam or an embankment that has failed to protect planned developments (7). In Malaysia, flash flood is not a strange phenomenon anymore. As a tropical climate, Malaysia has high humidity and rainfall throughout the year. Heavier rains usually fall during November to February, especially in the East Malaysia. Furthermore, the rapid urbanization has led to increase loss of open space and forested land, reducing flood plain and gradual canalization of the rivers. These factors have led to runoff and peak flow resulting in the increase of flood occurrences, especially in the city areas (8). In Malaysia, Department of Irrigation and Drainage (DID) is responsible for handling, managing, and reducing the flash flood. DID has identified and listed flash flood locations in each state by the order of severity. The severity is based on the flash flood frequency. Level 1 refers to cases when the frequency exceeds 5 times a year, level 2 refers to cases when the frequency is between 2-4 times a year, and finally, level 3 refers to cases when the flood happens less than 1 time during a year. Furthermore, the authors in (9) created the plotted map based on the estimation on road map published by the Malaysian Public Work Department (PWD) (see Figure 1). Three colours were used to differentiate flash flood duration.

As shown in Figure 1, Kuala Lumpur is exposed to most flash flood hazard in Malaysia. Kuala Lumpur is the capital city of Malaysia where several flash flood occurrences have been recorded over the past years. Between 1965 and 2016, 76 cases of natural disasters have been recorded in Malaysia, including landslide, epidemic, tsunami, mudflows, storm, and wildfire; half of the cases is related to flood. The impacts of the recorded flood include damage to properties, disruption in social activities and transportation systems, and losses in businesses and household assets. Concerned by the above-mentioned impacts, in 1971, Federal Government of Malaysia established a permanent commission for flood control to cope with the issue. The Head of the Commission is Minister of Agriculture, and DID has been given a mandate to take the flood mitigation initiatives (10). To identify the root causes of the flash flood, an Ishikawa diagram was constructed. In addition, 5W and 1H (5W - What, When, Where, Who and Why, 1H – How) approach has also been used to have a clear picture on how to tackle this issue as shown in Figure 2. Using the approach, related root cause analysis of flood situation in Kuala Lumpur can be divided into a few perspectives; Process, Technology, Economy, People and Social. Wet and humid climate as well as geographical characteristics of its location, Kuala Lumpur has a high potential to be exposed to flooding. Despite the existence of draining or water tunnels system in the city, several reservoirs and rivers have been found near the city, which will affect the water flow to the city.

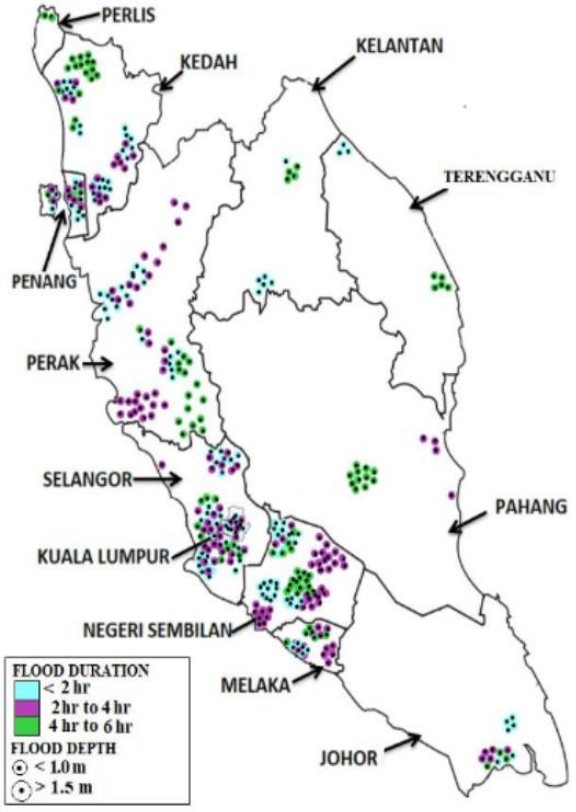


Figure 1: Flash Flood in Malaysia (source: (9))

2.2 Predictive Analytics for Early Flood Detection

Flood is a disaster Kuala Lumpur encounters almost every year. It often causes damage to belongings for business and people. In certain cases, there is a need to evacuate buildings in a certain area. At the end, the people request a vast amount of money for repairing and replacing necessities. Flash flood prediction is an effective solution that can reduce the impact of disaster. Predicting flash flood is quite challenging since the factors of prediction are not only depending on meteorological phenomena and heavy rain. Therefore, at the same time, it is very critical to apply flood forecasting as real-time indication of flow rates and water levels ahead of time.

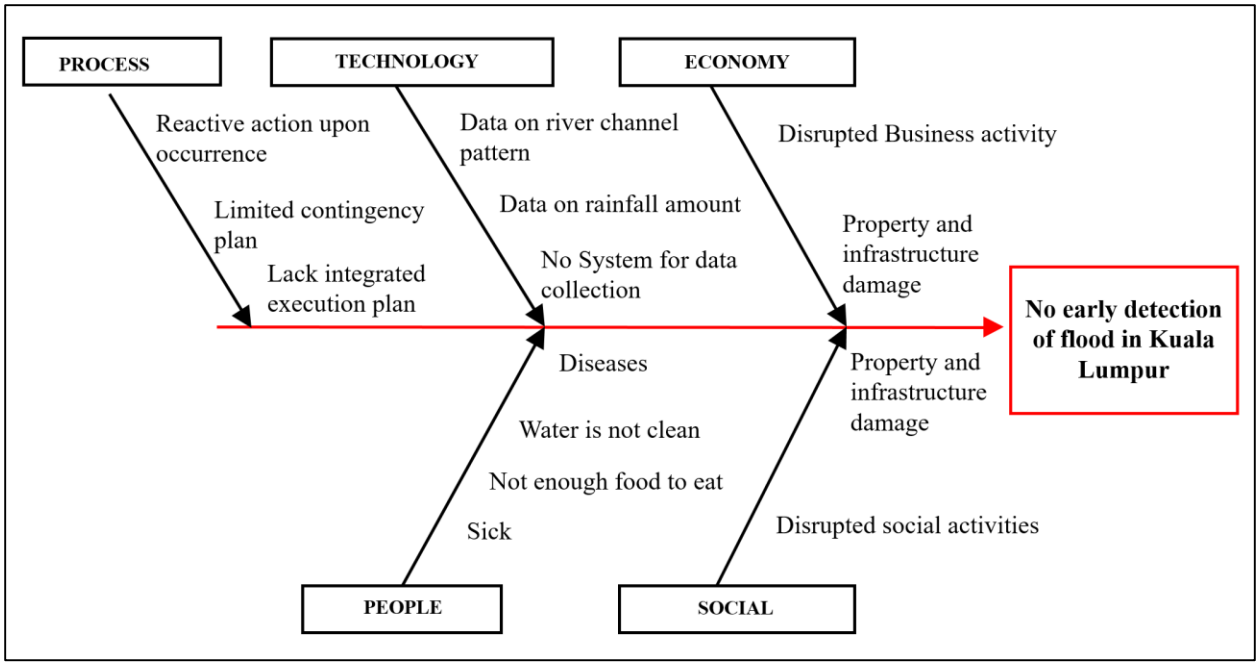


Figure 2: Ishikawa Diagram identifying the issues and factors of flash flood in Kuala Lumpur

Predictive analytics is one of the element of big data analytics. It is one of the technology categories defined in the elements of the fourth industrial revolution (4IR) and refers to a large collection of complex data generated by various instruments and sources. Big data analytics has the potential to be implemented by data-driven decision makers in organizations. To discover required data and to get information, data analysis approaches such as descriptive, predictive, and prescriptive are used. Therefore, with the result obtained from the analysis, data are transformed into predicted information to support better decision making. Predictive analytics will extract information from data and using it to predict trends and behaviour patterns. In relation to the issue of flash flood in city areas, this study will identify pattern within data to predict future flash floods and trends.

Commonly, statistical inferences, machine learning, artificial intelligence and data mining approaches are applied to predictive analytics. This approach has been found effective, especially when dealing with large datasets (11). Hence, it is essential to develop better understanding about what data can be used for flash flood prediction. By characterizing and knowing the data types, then the research will be able to identify which techniques and methods can be used for prediction. According to (11), there are three predictive analytics techniques known as regression, classification and clustering. Regression is a statistical technique to measure the relationship between variables. Linear and Non-Linear are two types of regression commonly uses for prediction. Meanwhile classification covers the process of categorizing a new instance based on information from a training dataset. By using the data attributes features from the training data set, the classifier can predict categories from unknown instances. Well-known classification methods include Bayesian, decision trees, Artificial Neural Networks and Support Vector Machine (SVM). Classification is similar to classification which also attempts to categorize a new observation into a class membership, only that this technique is an unsupervised method that discovers the natural groupings of a data set with unknown class labels.

It is essential to understand the flash flood data based on the prediction of the rainfall model. The research found the major factors of Kuala Lumpur's flash flood as follow: i) blockage of drainage channels, ii) increase of surface runoff, iii) uncontrolled urbanization, iv) poor drainage maintenance, and v) inadequate drainage system. Every flood event is followed by several physical and mental effects. Major physical effects of flash flood in Kuala Lumpur are damages to properties, business premises, and roads. As for the services, flash flood interrupts transportation services. According to [12], currently DID has come out with a list of mitigation actions such as: i) more frequent maintenance of drainage system, ii) construction of flood diversion, iii) control of land use activities, iii) installation of flood bypass, iv) limitation of deforestation, and v) enhanced and stringent acts on flash floods.

Based on literature review, flat clustering was used to categorize data and provide flood predictions. Flat clustering is generally used to limit classification of data. The analysed data is later transformed into dashboard, where the analysed data is presented in a visual form. These data give better insight and data understanding for the decision makers to predict the flood situation in an early time. As proposed by (12), inundation maps, which can be prepared in GIS, are useful tools to determine the areas vulnerable to flood events. The usage of geospatial data can help to construct Flood Hazard Map, Flood Risk Map, and Flood Evacuation Map in Malaysia (13). These maps are important to identify the risks involved and facilitate proper land use planning. Furthermore, the authors in (14) indicated that the usage of

satellite image data and geospatial data in flood forecasting processes leads to long-term risk reduction and effective post-disaster responses.

Advancement in tools employed to analyse complex data such as hydro climate data can accelerate the process of data computation (15). As such, the method of Nonlinear Autoregressive network with Exogenous input (NNARX) uses numerical method to examine simulated and actual data. Through examinations in real situations, NNARX model was found capable of predicting flood water level ahead of time (16). Meanwhile, another study took into consideration the non-linear least square regression as a method applicable to the estimation of the rain intensity by measuring the size of rain drops (17). The study on the correlation of flood pattern (water level) and the prediction (rainfall) on its recursive occurrence is of a high significance (18).

The rapid advancement of technology of Big Data Analytics (BDA) can be used to help the early prediction of flood. Study from (19, 20) indicated that, water level prediction has improved from 5 hours of early flood prediction to 3 hours using the same algorithm Nonlinear Autoregressive network with Exogenous input (NNARX). Furthermore, by using advance data analysis from collection of live heterogeneous data sources like sensors and satellite, process of prediction can be done in timely manner, an improvement from what is lacking in the previous generation of data analysis environment. Therefore, providing timely information helps the related parties to take further actions and implement prepared strategic plans.

The early flood detection and its visualization need to be improved in future studies carried out in this field. The gaps are identified as follow:

1) The data visualization dashboard can be understood by the decision makers and operators (people/society); therefore, data presentation should be clearly expressed and it should satisfy the requirements of decision makers and operators. The artificial intelligence applied to the system can be useful in improving the model. Thus, this will increase the participation of both government and society in taking proactive action regarding early detection of flood preventing its damages.

2) The early flood detection model based on real-time data is suggested, whereby it can provide the government and society with more accurate information, improving decision making and execution plans. Thus, the flood-induced damages can be reduced and economy will be secured.

3) The data analysis of early flood detection model should include other variables rather than focusing only on rainfall data and water level. With the other variables, it can expand the analysis widely with more data pattern.

2.3. The benefit of flash flood prediction

Literature comprises multiple variables, including demographic, hydrological, and ecological variables, that have the potential to enhance the decision making capabilities in reducing water-based disaster risk (2, 21). Early detection result of analytic data will help the users or stakeholders to make better flood-related decisions such as town planning, flood prevention program, identifying strategies and actions in order to reduce damages. Most importantly, prediction is the catalyst for flood mitigation plan. The presentation of the flood risk prediction will be simplified using data visualization in order to improve the users' understanding about the data and make use of the information.

Proper planning and mitigation at the identified area can help to reduce the local economy losses, thus, economy in the high-risk flood area will be more sustainable. Land and property value can be appreciated when the flood can be predicted and the risk of land disruption, landslide, and cracks can be reduced. As such, the town planning can avoid massive development in the high-risk flood area. This action can help to preserve the land and greens in the city that lead to more secure sustainable environment.

The awareness and alertness about the predictive flash flood will help the people to make a proper planed about the activities during the rainy season. The social and business activities can operate as usual without worrying the risk caused by the flood. Furthermore, the people can customize their activities according to the season as predicted by the result. Daily activities and events can be held with ease without any disruption and sudden cancellation. In the other hand, people are more alert and concern about their safety surroundings, thus to be more responsible to prevent the any risk and make a better living.

3 Methodology

This research is carried out using the Cross-Industry Standard Process for Data Mining known as CRISP-DM. CRISP-DM is a widely-used methodology and its use in projects helps researchers to be focused on delivering real business value. CRISP-DM breaks the process of data mining into six major phases as shown in Figure 2. The sequence of the phases is not strict and moving back and forth between different phases is allowed, as it is always required. The arrows in the process diagram indicate the most important and frequent dependencies among phases. The outer circle in the diagram symbolizes the cyclic nature of data mining itself.

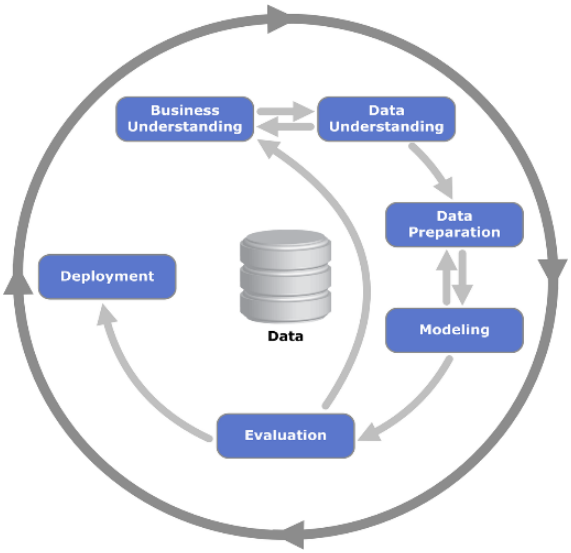


Figure 3: CRISP-DM Cycle

This paper will describe only the data understanding phase since it is our main focus in this study. The data understanding phase involves four steps: i) Initial Data Collection, ii) Describing data, iii) Exploring data, and iv) Verifying data quality. Furthermore, this paper will go deeper only regarding the initial data collection step. Initial data collection is essential since it is the main step towards getting the knowledge about the data, the needs that the data will satisfy, the availability, requirements, and the sources. This step is pre-requisite before real data collection because real world data is often inadequate, unreliable, and/or missing in certain behaviours or trends and is likely to contain many errors. Therefore, the characterization of the data in the initial step is capable to collect data from different sources and then understand, re-organize, re-format, and integrate the collected data (5) before the real data collection takes place. It requires an extensive analysis to deal with data scalability based on the selection of data requirement. Moreover, data characterization also concerns more about the business process and the flow of data within it. There are lots of data and knowledge that can be extracted from business processes and from relationship among the data. Finally, proven and credible initial data understanding is significant to determine the quality and trust of the analysis results (22).

4 Flash Flood Initial Data Understanding

Since flash flood takes place due to multiple factors, it is difficult to understand data involved and choose the right parameters unique to the impacted area. For example, coastal areas and vegetation areas versus a place that is nearby a river with rapid development may have different factors to be considered. Thus, any single model cannot offer all factors from all aspects when looking into the best way of understanding the data for later analysis.

In Malaysia, there is two types of flood: monsoon flood and flash flood (or urban flood). In literature, several contributing factors for flood such as tidal rivers, climate change, temperature, altitude, drainage, and terrain changes have been discussed (23). Since this paper is focused on controlling the flash flood in Malaysia, only related notable key variables of this type of flood are discussed in the following. Principally, flash flood occurs due to four major factors: rainfall, blocked drainage, urbanization, and fluvial flooding (24). Each factor will be described in the following subsections.

4.1 Rainfall

A study into the historical data of rainfall indicates that extreme rainfall is associated with flash floods or water river floods. It is clear that the key factor in river flash severity and its speed is characterised by intensity of the rainfall; thus, it will cause risk to people's life (25). The authors in (26, 27) used rainfall data, which were associated with the hydrological data, to resolve flood disaster using an early warning system. Another study (28) showed that a significant relation exists between rainfall and the disaster mapping in metropolitan and surroundings area. For example, in Philippines, a predictive modelling technique, i.e., the clustering technique, is used for flood projection and warnings by the use of rainfall information through web applications and establishing the information on alternate routes (29). Thus, with the analysis of the information about rainfall and its regime, the results can help the operators know the current and historical rainfall water information promptly and accurately and also help policy makers design the right flood rescue programs, hence minimizing casualties and property losses (27).

Thus, rainfall data can be used for prediction purposes. The researchers believe that the fluvial flood data that has been collected during 10 years (from 2000 to 2010) by DID can bring insight into predicting the flash flood in city areas. However, the research also found that rainfall data alone are insufficient to determine flash flood. By analysing trend using time series and historical data, the correlation result showed a weak relationship between rainfall amount and number of flood incidents. This study found that flood incidents do not happen just because of rainfall; there are also other contributing factors such as tidal rivers, climate change, temperature, altitude, drainage and terrain changes (23). Due to the focus on flash flood, the factors such as tidal rivers, temperature and climate change will not be covered in this study. Thus, the research will further investigate the factors of altitude, drainage and terrain changes in the next paragraph.

Table 2: Maintenance and Number of Reservoirs in Malaysia

State	DID	PBT	Others	Total	Area (ha)
Perlis	35	18	0	53	96.55
Kedah	23	138	8	169	233.81
Penang	21	185	34	240	38.00
Perak	10	278	0	288	214.40
Selangor	38	188	171	397	1,492.92
Kuala Lumpur	13	1	0	14	305.03
Putrajaya	0	10	0	10	5.76
Negeri Sembilan	7	111	93	211	438.38
Melaka	0	6	130	136	152.12
Johor	15	302	33	350	253.95
Pahang	13	11	21	45	221.64
Terengganu	9	33	9	51	412.39
Kelantan	1	4	24	29	14.63
Sarawak	14	172	0	186	54.17
Sabah	0	21	8	29	70.43
Labuan	1	2	0	3	1.68
Total	**200**	**1,480**	**531**	**2,211**	**4,005.85**

Source: DID (Department of Irrigation & Drainage)
PBT (Pihak Berkuasa Tempatan means Local Authorities in Malaysia)

4.2 Blocked Drainage

Drainage is the natural or artificial removal of surface

and sub-surface water from an area. Blocked drainage is often caused by debris and litter, particularly around the cities. According to The Star (30), silt traps built is very poorly maintained; as a result, sediment and debris end up clogging the drains, leading to flash flood after a downpour. For years, the media has been highlighting the poor maintenance of the city's drains that have caused reduced flow capacity and, thereby causing flood in the area. Data related to drainage are collected by KL Municipal Council or DBKL. DBKL has started to collect flash flood-related data only since 2010.

4.3 Urbanization

Urbanization is the outflow process of population moving to a larger city in order to satisfy their needs for education, cultural development and professional realization (31). In (32), a very close relationship was found between flooding and rapid development and urbanization. The more populous the area, the higher the chances of flash flood to occur (33). The Geographical Information System (GIS)-based analysis indicates that there is a strong correlation between the land use changes and the flood hazards (34). If development processes are not done in a sustainable approach, it would result in higher likelihood for flood occurrences. The Integrated Flood Forecasting and River Monitoring System (IFFRM) has been developed by DID in order to forecast flood by monitoring the river within Klang Valley.

The research found strong correlation between flash flood and urbanization because of the development of urban area changes the infrastructure and its surrounding. The changes in land use associated with urbanization affect flooding in many ways. Land use and massive human activities influence and modify how rainfall are stored on and run off the land surface into rivers. In cities areas which covered more by roads and buildings, the land surfaces have less capacity to store the rainfall. With less storage capacity for water in urban area, urban streams rise more quickly during heavy rain and have higher density of sediments. Indirectly, this impacted the drainage system which eventually leads debris and sediments to the river channel.

4.4 Fluvial Flooding

Areas prone to flood is another contributing factor of flood occurrence, especially areas situated at the lower height, adjacent to a river called floodplain(35). Kuala Lumpur has its own fair share of flood history. Being situated at the confluence of Kelang and Gombak rivers has added another factor of flash flood occurrence by having riverine or fluvial flood in the area. DID has installed telemetry stations to monitor river water level at the rivers in Malaysia as part of the Integrated River Basin Management (IRBM) Plan and Flood Mitigation Plan to reduce flood risk of the area (36). Apart from that, telemetry stations are also used to monitor rainfall intensity daily. To complete the monitoring component, IP cameras were also installed to have a better picture of the condition.

Siren will be used once the water level reaches the warning and danger level. The siren is only to warn nearby residents and is stationed at certain places with high track records of flooding.

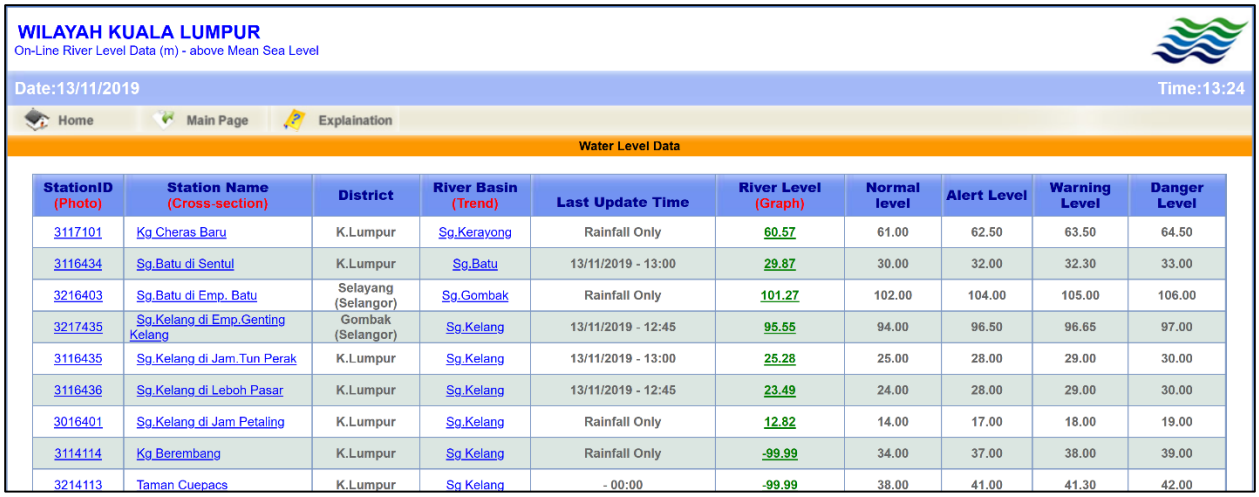
Figure 4: Warning Level Indicator

In Selangor alone, there are about 66 water level telemetry stations, 95 rainfall intensity telemetry stations, 103 siren telemetry stations, 30 IP cameras telemetry stations, and 25 flood gauge telemetry stations, as part of flood early warning and forecasting system installed by DID. These data and information will then be fed into a portal for monitoring (37).

Figure 5: Flood Gauge

5 Conclusion

The flood situation in the city of Kuala Lumpur contributes major impact to the society, economy, and environment. Climate change is one of the main concerns that contribute to flooding. The implementation of big data analytics approaches such as predictive and descriptive analyses will enable users to identify the correlation between rainfall and water level and data patterns, which result in early prediction for flood. If prediction and preventive actions are done properly, the damage to people and environment can be significantly reduced.



WILAYAH KUALA LUMPUR
On-Line River Level Data (m) - above Mean Sea Level

Date:13/11/2019 Time:13:24

Home | Main Page | Explaination

Water Level Data

StationID (Photo)	Station Name (Cross-section)	District	River Basin (Trend)	Last Update Time	River Level (Graph)	Normal level	Alert Level	Warning Level	Danger Level
3117101	Kg Cheras Baru	K.Lumpur	Sg.Kerayong	Rainfall Only	60.57	61.00	62.50	63.50	64.50
3116434	Sg.Batu di Sentul	K.Lumpur	Sg.Batu	13/11/2019 - 13:00	29.87	30.00	32.00	32.30	33.00
3216403	Sg.Batu di Emp. Batu	Selayang (Selangor)	Sg.Gombak	Rainfall Only	101.27	102.00	104.00	105.00	106.00
3217435	Sg.Kelang di Emp.Genting Kelang	Gombak (Selangor)	Sg.Kelang	13/11/2019 - 12:45	95.55	94.00	96.50	96.65	97.00
3116435	Sg.Kelang di Jam.Tun Perak	K.Lumpur	Sg.Kelang	13/11/2019 - 13:00	25.28	25.00	28.00	29.00	30.00
3116436	Sg.Kelang di Leboh Pasar	K.Lumpur	Sg.Kelang	13/11/2019 - 12:45	23.49	24.00	28.00	29.00	30.00
3016401	Sg.Kelang di Jam Petaling	K.Lumpur	Sg.Kelang	Rainfall Only	12.82	14.00	17.00	18.00	19.00
3114114	Kg Berembang	K.Lumpur	Sg Kelang	Rainfall Only	-99.99	34.00	37.00	38.00	39.00
3214113	Taman Cuepacs	K.Lumpur	Sg Kelang	- 00:00	-99.99	38.00	41.00	41.30	42.00

Figure 6: Infobanjir Portal showing Water Level & Rainfall Data

Flash flood pattern prediction is still in the early stages of development. The evaluation of interdependent variables as contributing factors is crucial in determining the correlations among different variables. In Malaysia, Big Data Analytics has been started since 2014 mandate. From that time, data collections conducted by DID on river basins in regard to water level have been a big help given to researchers to make required analyses on time series and historical data collected. Collaboration on data sharing like this has helped researchers a lot in conducting prediction tasks using machine learning algorithms. This indicates that collaboration between different parties such as government agencies, researchers, and private sectors can accelerate the formulation of predictive models. In future, prediction models can be completed and then can be used as a breakthrough in predicting flash flood patterns in Kuala Lumpur.

The method that integrate structural and non-structural measures can manage the issues in a more realistic and holistic way. Therefore, integration of various parties can construct a better framework to overcome flash flood in Malaysia and provide sustainable development environment for the surrounding area.

Four flash flood incidence factors identified in the previous section are significant clues for real data collection. The characterization of the flash flood factors will act as the guidelines to identify the potential data collection that can be used to predict flash flood in Kuala Lumpur, Malaysia. Furthermore, this research will collect the identified factors from the agencies, describe and explore the data, and finally it will verify again the data quality.

Acknowledgement

The authors gratefully acknowledge Ministry of Education (MOE) and Universiti Teknologi Malaysia (UTM) for the support in conducting this study. This work is conducted at Razak Faculty of Technology and Informatics and funded by UTM (RUG: QK130000.2656.17J23).

Ethical issue

Authors are aware of, and comply with, best practice in publication ethics specifically with regard to authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests and compliance with policies on research ethics. Authors adhere to publication requirements that submitted work is original and has not been published elsewhere in any language.

Competing interests

The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.

Authors' contribution

All authors of this study have a complete contribution for data collection, data analyses and manuscript writing

References

1. Service NsNW. Flash Flooding Definition 2019 [Available from: https://www.weather.gov/phi/FlashFloodingDefinition.
2. Zambrano L, Pacheco-Munoz R, Fernandez T. Influence of solid waste and topography on urban floods: The case of Mexico City. Ambio. 2018;47(7):771-80.
3. Malik A, Maciejewski R, Towers S, McCullough S, Ebert DS. Proactive Spatiotemporal Resource Allocation and Predictive Visual Analytics for Community Policing and Law Enforcement. IEEE Transactions on Visualization and Computer Graphics. 2014;20(12):1863-72.
4. Attaran M, Attaran S. Opportunities and Challenges of Implementing Predictive Analytics for Competitive Advantage. International Journal of Business Intelligence Research. 2018;9:1-26.
5. Junhua LU WC, Yuxin MA,Junming KE,Zongzhuang LI,Fan ZHANG,Ross MACIEJEWSKI. Recent progress and trends in predictive visual analytics. Front Comput Sci. 2017;11(2):192-207.
6. Lebied M. Top 11 Business Intelligence and Analytics Trends for 2017 2016 [Available from: http://www.datapine.com/blog/business-intelligence-trends-2017/.

7. Jha AK, Bloch R, Lamond J. Cities and Flooding A Guide to Integrated Urban Flood Risk Management for 21st Century. 2012.
8. Seang SH. A Case Study of Mitigating Flood in City Center of Kuala Lumpur. 2009.
9. Nizam MS, Ghani ANA, Md Ghazaly Z. The characteristics of road inundation during flooding events in peninsular Malaysia. International Journal of GEOMATE. 2019;16:129-33.
10. Loi HK. Flood Mitigation and Flood Risk Management in Malaysia. 1996.
11. Assunção MD, Calheiros RN, Bianchi S, Netto MAS, Buyya R. Big Data computing and clouds: Trends and future directions. Journal of Parallel and Distributed Computing. 2015;79-80:3-15.
12. Gogoase DEN, Armaş I, Ionescu CS. Inundation Maps for Extreme Flood Events at the Mouth of the Danube River. International Journal of Geosciences. 2011;02(01):68-74.
13. Zakaria SF, Zin RM, Mohamad I, Balubaid S, Mydin SH, Mdr EMR. The development of flood map in Malaysia. 2017.
14. Yu M, Yang C, Li Y. Big Data in Natural Disaster Management: A Review. Geosciences. 2018;8(5).
15. Abdullah MF, Ibrahim M, Zulkifli H. Big Data Analytics Framework for Natural Disaster Management in Malaysia. Proceedings of the 2nd International Conference on Internet of Things, Big Data and Security2017. p. 406-11.
16. Rohaimi NA, Ruslan FA, Adnan R, editors. 3 Hours ahead of time flood water level prediction using NNARX structure: Case study pahang. 2016 7th IEEE Control and System Graduate Research Colloquium (ICSGRC); 2016 8-8 Aug. 2016.
17. Reba MNM, Roslan N, Syafiuddin A, Hashim M, editors. Evaluation of empirical radar rainfall model during the massive flood in Malaysia. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS); 2016 10-15 July 2016.
18. Yusoff A, Din N, Yussof S, Khan S. Big data analytics for Flood Information Management in Kelantan, Malaysia2015. 311-6 p.
19. Ruslan FA, Samad AM, Adnan R. 3 Hours Flood Water Level Prediction Using NNARX Structure: Case Study Kuala Lumpur 2015.
20. Adnan R, Samad AM, Zain ZM, Ruslan FA. 5 hours flood prediction modeling using improved NNARX structure: case study Kuala Lumpur. 2014 IEEE 4th International Conference on System Engineering and Technology (ICSET)2014. p. 1-5.
21. Youssef AM, Sefry SA, Pradhan B, Alfadail EA. Analysis on causes of flash flood in Jeddah city (Kingdom of Saudi Arabia) of 2009 and 2011 using multi-sensor remote sensing data and GIS. Geomatics, Natural Hazards and Risk. 2015;7(3):1018-42.
22. Sacha D, Stoffel A, Stoffel F, Kwon BC, Ellis G, Keim D. Knowledge Generation Model for Visual Analytics. Visualization and Computer Graphics, IEEE Transactions on. 2014;20:1604-13.
23. Mohamed NH, Ismail A, Ismail Z, Adnan CWMSW, Raji MFA. TREND ANALYSIS AND FORECASTING OF RAINFALL AND FLOODS. 2014.
24. Tariqur Rahman Bhuiyan, Mohammad Imam Hasan Reza, Er Ah Choy, Pereira JJ. Direct Impact of Flash Floods in Kuala Lumpur City. ASM Sci J, 11(3), 145-157. 2018.
25. Archer DR, Fowler HJ. Characterising flash flood response to intense rainfall and impacts using historical information and gauged data in Britain. Journal of Flood Risk Management. 2018;11(S1):S121-S33.
26. Garcia FCC, Retamar AE, Javier JC. Development of a Predictive Model for On-Demand Remote River Level Nowcasting: Case Study in Cagayan River Basin, Philippines. 2016 IEEE Region 10 Conference (TENCON); Singapore, Singapore IEEE; 2016.
27. Liu X, Hao L, Liu Y, editors. The Research on Flood Control and Rainfall Regimen Information System of Poyang Lake Areas. 2009 International Conference on Environmental Science and Information Application Technology; 2009 4-5 July 2009.
28. Althuwaynee O, Pradhan B, Ahmad N. Estimation of rainfall threshold and its use in landslide hazard mapping of Kuala Lumpur metropolitan and surrounding areas. Landslides. 2015;Vol 12(No 5):pg.861-75.
29. Panganiban EB, Cruz JCD, editors. Rain water level information with flood warning system using flat clustering predictive technique. TENCON 2017 - 2017 IEEE Region 10 Conference; 2017 5-8 Nov. 2017.
30. Lam Thye L. Improve city's drainage2016. Available from: https://www.thestar.com.my/opinion/letters/2016/05/16/improve-citys-drainage/.
31. Yuliya A.Efremova, Olga N.Goryacheva , F.Kurbanova R. Image of a City as a Factor of Strategic Development of a Territory. Journal of Environmental Treatment Techniques. 2019;7(Special Issue on Environment, Management and Economy):925-9.
32. Siti Fadzilatulhusni Mohd Sani, Rindam M. Analisis taburan hujan dan impaknya kepada sumber air di Pulau Pinang. 2011.
33. Bhuiyan Tariqur R. Facts and Trends of Urban Exposure to Flash Flood: A Case of Kuala Lumpur City. In: Reza Mohammad Imam H, editor. Improving Flood Management, Prediction and Monitoring. Community, Environment and Disaster Risk Management. 20: Emerald Publishing Limited; 2018. p. 79-90.
34. Chang H, Franczyk J, Kim C. What is responsible for increasing flood risks? The case of Gangwon Province, Korea. Natural Hazards. 2009;48(3):339-54.
35. Mohammad Ali Nezammahalleh, Mojtaba Yamani, Abolghassem Goorabi, Mehran Maghsoudi, Mohamadkhan S. Evaluation of a GIS-based floodplain height difference model for flood inundation mapping, case study Rudbar, Iran. Journal of Environmental Treatment Techniques. 2017;5(3):100-6.
36. DID DoIaD. Towards Realising Integrated River Basin Management In Malaysia 2017 [Available from: https://www.water.gov.my/index.php/pages/view/708.
37. Selangor JN. Pengurusan Sumber Air dan Hidrologi JPS Negeri Selangor 2019 [Available from: http://water.selangor.gov.my/index.php/ms/2-uncategorised/84-pengurusan-sumber-air-dan-hidrologi-jps-negeri-selangor.
36. Deshpande, P. (2018). Predictive and prescriptive analytics in big data Era. Advances in Intelligent Systems and Computing, 810, 123–132.
37. Seebacher, D., Miller, M., Polk, T., Fuchs, J., & Keim, D. (2019). Visual Analytics of Volunteered Geographic Information: Detection and Investigation of Urban Heat Islands. IEEE Computer Graphics and Applications.
38. Mukherjee, S. (2019). Predictive Analytics and Predictive Modeling in Healthcare. SSRN Electronic Journal, (June).
39. Yeon, H., Kim, S., & Jang, Y. (2017). Predictive visual analytics of event evolution for user-created context. Journal of Visualization, 20(3), 471–486.
40. Cashman, D., Humayoun, S. R., Heimerl, F., Park, K., Das, S., Thompson, J., … Chang, R. (2019). A User-based Visual Analytics Workflow for Exploratory Model Analysis. Computer Graphics Forum, 38(3), 185–199.